

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Tomaž Tomažič

**NAPOVEDOVANJE DNEVNE
PROIZVODNJE ELEKTRIČNE
ENERGIJE SONČNIH
ELEKTRARN**

MAGISTRSKO DELO
ŠTUDIJSKI PROGRAM DRUGE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: izr. prof. dr. Matjaž Kukar

Ljubljana, 2016

Rezultati magistrskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov magistrskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

Zahvaljujem se mentorju izr. prof. dr. Matjažu Kukarju za mentorstvo, strokovno pomoč, nasvete in usmeritve pri izdelavi magistrskega dela.

Zahvaljujem se tudi Agenciji Republike Slovenije za okolje (ARSO) za vremenske napovedi in Nevi Pristov in Juretu Cedilniku za strokovno razlago pri razumevanju teh podatkov.

Zahvaljujem se tudi podjetju Elektro Primorska za podatke o proizvodnji in hčerinskemu podjetju E3, še posebej pa Juriju Juršetu, Marku Vetrihu, Boštjanu Pičulinu, Dušanu Tomažiču in Davidu Ferjančiču za pomoč pri pridobitvi in razumevanju podatkov sončnih elektrarn.

Kazalo

Povzetek

Abstract

1	Uvod	1
1.1	Cilji in motivacija	5
1.2	Metodologija	9
1.3	Prispevek naloge	10
2	Pregled področja	13
2.1	Raziskave v Sloveniji	13
2.2	Tuje raziskave	15
3	Podatki	19
3.1	Proizvodnja elektrarn	21
3.2	Vremenske napovedi	35
4	Napovedovanje	41
4.1	Izračun atributov	41
4.2	Modeliranje proizvodnje	49
5	Rezultati	55
5.1	Sklepne ugotovitve	56

KAZALO

6 Zaključek	63
6.1 Nadaljne delo	64
A Podrobnejši rezultati napovedi	75

Seznam uporabljenih kratic

kratica	angleško	slovensko
CRISP-DM	Cross Industry Standard Process for Data Mining	Standardni proces podatkovnega rudarjenja
ALADIN	Aire Limitée, Adaptation Dynamique, Développement International	Omejeno območje - dinamična adaptacija, mednarodno sodelovanje
GB	giga byte	giga bajt
TP	conversion station	transformatorska postaja
PV	photovoltaic	fotonapetostni
NOCT	Nominal Operating Cell Temperature	predpisana delovna temperatura celice
STC	Standard test conditions	standardni test razmer
ARIMA	auto-regressive integrated moving average	auto-regresijsko integrirano tekoče povprečenje
ARMA	auto-regressive moving average	auto-regresivno tekoče povprečenje
SAIFI	system average interruption frequency index	povprečni faktor sistemskih izpadov
STO	transmission system operator	sistemski operater prenosnega elektroenergetskega omrežja
SODO	electricity distribution system operator	sistemski operater distribucijskega omrežja
OVE	a Renewable energy sources	obnovljivi viri energije
WGS84	Earth-centered, Earth-fixed terrestrial reference system and geodetic datum.	Eden izmed standardnih sistemov po katerem lahko določamo lokacijo.

Povzetek

Naslov: Napovedovanje dnevne proizvodnje električne energije sončnih elektrarn

Slovenija je v zadnjih letih doživela precejšen razmah električne proizvodnje iz obnovljivih virov energije, med katerimi prevladuje sončna energija, saj cene fotonapetostnih modulov sčasoma strmo padajo in se vedno več ljudi odloča za takšne investicije. Napovedovanje proizvodnje električne energije iz fotonapetostnih modulov je pomembno tako za elektro distributerje, kot za trgovalce na borzi.

V magistrskem delu pokažemo različne pristope obdelave surovih podatkov, pridobljenih iz merilnih mest in njihove vizualizacije, ki so zelo pomembne za lažje razumevanje podatkov. Attribute na podlagi katerih napovedujemo, najpogostejše in najenostavnejše pridobimo v obliki modelskih napovedi vremenskih parametrov. Tisti, ki so ocenjeni kot najkoristnejši, so uporabljeni za napovedovanje proizvodnje električne energije z različnimi modeli strojnega učenja. Raziskano je tudi vključevanje vpliva več različnih modelskih točk na določeno lokacijo elektrarne. Napovedovanje je prilagojeno primorski regiji Slovenije. Naši rezultati so v primerjavi z različnimi sorodnimi deli primerljivi ali celo boljši.

Ključne besede: podatkovno rudarjenje, sončne elektrarne, dnevno napovedovanje, podatkovni tokovi.

Abstract

Title: Prediction of daily photovoltaic systems production

Slovenia has vastly expanded electricity production from renewable energy sources recently. In the renewables world solar energy prevails, because in last few years prices of photovoltaic modules have fallen steeply, which gives people extra motivation to invest in photovoltaic systems. Predicting electricity production from photovoltaic modules is very important for electricity distributors and traders in electric energy markets. We describe the Slovenian electric energy market with focus on daily products in which our predictive model can be applied.

In the thesis we show different approaches of processing raw data given from power plants and its visualisations, which are very important for easier understanding of the data. Attributes which are used for predictions are usually obtained in the form of weather forecast model parameters. Only the most valuable attributes are used in different machine learning models for predicting electricity production. Influence of spatial averaging multiple weather predictions for every power plant separately are studied, but our predictions are adjusted for Primorska region of Slovenia. We discuss and compare our results with other recent researches, where we reached a comparable or even better results.

Keywords: data mining, solar power stations, daily forecast, data streams.

Poglavje 1

Uvod

Obnovljivi viri energije (OVE), kot so sončna, vetrna in vodna energija, vzbujajo vedno večjo pozornost, predvsem zaradi manjšega negativnega vpliva na okolje v primerjavi s fosilnimi gorivi. Cena obnovljivih virov je še vedno visoka, predvsem zaradi težke integracije v obstoječe električno omrežje. Razlog za to je v težavnem napovedovanju proizvodnje in posledično kompleksnejšemu sistemu za izravnavanje proizvodnje in odjema, vendar se ta cena vsako leto znižuje. Z naprednimi razsmerniki omogočajo OVE, poleg prilagajanja delovne moči, tudi prilagajanje jalove moči za regulacijo nape-tosti v distribucijskem omrežju, zaradi česar se lahko poveča tudi skupna priključna moč omrežja. Tako se torej delno izognemo dragim investicijam v povečevanje zmogljivosti distribucijskega omrežja, kar naj bi se sčasoma poznalo tudi z nižjo omeržnino vsakega gospodinjstva. Sončna energija je eden izmed trenutno najbolj uporabljanih obnovljivih virov energije, saj cene fotonapetostnih modulov strmo padajo. Pretvorba sončne energije s pomočjo fotonapetostnih modulov predstavlja eno izmed možnosti uporabe in pretvorbe sončne energije. Električno energijo lahko neposredno proizvajamo iz fotonapetostnih (PV) modulov povsod, kjer so primerni pogoji za postavitve in kjer je sončne energije dovolj [1].

Slovenija je v zadnjih letih doživela precejšen razmah proizvodnje iz obnovljivih virov energije. Pospešeno vključevanje tako imenovanih razpršenih

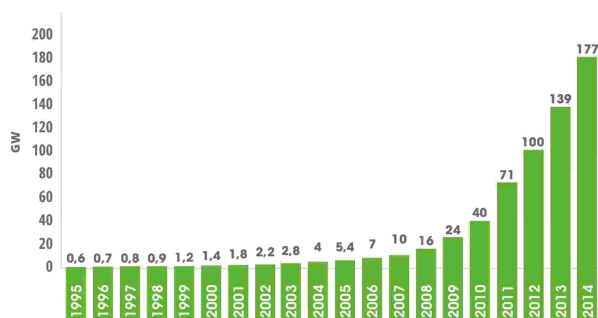
virov proizvodnje električne energije sovpada z vzpostavitvijo prenovljenega sistema finančnih spodbud za tovrstno proizvodnjo ob koncu leta 2009. Namen podporne sheme leta 2012 in 2013 za proizvodnjo iz obnovljivih virov energije [2] je bilo spodbujanje različnih tehnologij za izrabo obnovljivih virov energije, pri čemer so največjo porast doživele sončne elektrarne (slika 1.1).

Pri sončnih elektrarnah obstaja več dejavnikov, ki vplivajo na proizvodnjo fotonapetostnih modulov: difuzna svetloba, megla, prehodne oblačnosti, koncentracija prahu in umazanija na panelih, senčenja, temperatura, postopno padanje učinkovitosti modulov, sončni kot, okvare ... Vse naštetu učinkuje na količino ter hitro spremenljivo dinamiko oddane energije v omrežje. Pri napovedovanju proizvodnje iz razpršenih virov je veliko pozornosti potrebno posvečati tudi izrednim dogodkom, ki morda v preteklosti niso bili zaznani kot dejavniki, ki bi bistveno vplivali na elektroenergetski sistem in njegovo stabilnost. Lep primer je delni sončev mrk, ki se je zgodil 20. marca 2015. Slednji je terjal številne priprave in ukrepe ter spravil v visoko stanje pripravljenost večji del evropskih operaterjev prenosnih omrežji. Proizvodnja večine obnovljivih virov energije je torej po svoji naravi nestalna, hitro spremenljiva ter geografsko razpršena, s tem pa slabše vodljiva v primerjavi s proizvodnjo iz drugih (konvencionalnih) virov [3, 4]. Ker evropske smernice z leti zahtevajo večji delež (27% do leta 2030) pridobljene energije iz obnovljivih virov, in trend kaže da bomo dosegli 40%, postaja ta problem tudi vedno bolj aktualen [5]. V tabeli 1.1 je prikazana statistika moči in števila sončnih elektrarn v zadnjih letih. Konec leta 2013 vidimo vpliv uvedbe, s katero je slovenska vlada močno znižala odkupne cene električne energije, proizvedene iz sončnih elektrarn. To je njihovo rast zavrlo in jo skoraj popolnoma ustavilo. Skupna proizvedena energija iz sončnih elektrarn predstavlja $\sim 1\%$ celotne proizvodnje v Sloveniji.

Električna energija ima značilnost, da se porablja med proizvodnjo z malo ali celo brez prostora za shranjevanje. Električna se proizvaja glede na zahteve porabnikov iz mest, industrije in komercialnih odjemalcev. Veliki od-

Leto	2009	2010	2011	2012	2013	2014	2015
Število elektrarn	162	406	981	2425	3233	3322	3339
Moč [MW]	4,6	15,7	57,5	173,8	247	252,6	256,8
Povprečna moč [kW]	28	39	59	72	76	76	77

Tabela 1.1: Razvoj sončnih elektrarn v Sloveniji [6].



Slika 1.1: Svetovna nazivna moč fotovoltaičnih elektrarn [3].

jemalci morajo napovedati celoletni odjem, za vsako uro. Glavni problem vključevanja obnovljive energije v električno omrežje je njena nepredvidljivost. Pri koriščenju sončne energije za pridobivanje električne energije, nastopajo različne ovire. Glavni razlogi, zakaj se sončne energije ne izrablja več, so majhna gostota toka energije sevanja Sonca, slab izkoristek naprav, ki izrabljajo sončno energijo (sončne celice, sončni kolektorji) in nestabilno sevanje v uporabnem časovnem obdobju. Nepredvidljivost proizvodnje povzroča velika nihanja v omrežju. Da se ta nihanja zgladi, se uporablja različne metode. Leta 2012 so v Sloveniji prvič vzpostavili sistem virtualnih elektrarn¹ [7], problem se pa rešuje tudi s črpanjem vode v klanec, baterijami, agregati in podobnimi metodami [9]. Trenutno stanje v Sloveniji ni kritično za okvare, ker imamo veliko hidroelektrarn in v primeru prednostnega dispečiranja² s preveliko proizvodnjo, lahko hidroelektrarne spuščajo vodo mimo turbin ali pa črpajo vodo v višje ležeče rezervarje (HE Avče) ter tako zmanjšajo proizvodnjo oziroma povečajo porabo.

¹ Virtualna elektrarna predstavlja enega glavnih gradnikov pametnega elektroenergetskega omrežja prihodnosti. Tvori jo napreden računalniški sistem, ki omogoča dobavitelju električne energije učinkovito upravljanje z vedno številčnejšimi in nepredvidljivimi obnovljivimi viri, hranilniki energije in prilagodljivim odjemom. Gre za agregacijo različnih vrst elektro energijskih virov, ki lahko s svojo kapaciteto učinkovito nadomeščajo oziroma dopolnjujejo tradicionalne elektrarne. (povzeto po [8])

² Prednostno dispečiranje zagotavlja upravljalec prenosnega in upravljalci distribucijskega omrežja v skladu s pravili in navodili organizatorja trga. V skladu z energetskega zakonom in predpisi, mora organizator trga pri izravnavanju ponudb dajati prednost ponudbam kvalificiranih proizvajalcev električne energije. Kvalificirani proizvajalci so v prvi vrsti elektrarne, ki proizvajajo energijo iz vetra ali sonca, geotermalne energije ali biomase, torej iz obnovljivih virov (OVE) in v soproizvodnji toplote in električne energije z visokim izkoristkom (SPTE). Cena odkupa električne energije iz vseh kvalificiranih elektrarn določi Vlada Republike Slovenije. Cena odkupa je razčlenjena glede na dnevna in letna obdobja in je lahko različna za različne vrste kvalificiranih elektrarn ter starost elektrarn. V trgovanju se ne sme posegati na način izkrivljanja konkurence na trgu in diskriminatornosti članov. (povzeto po [10])

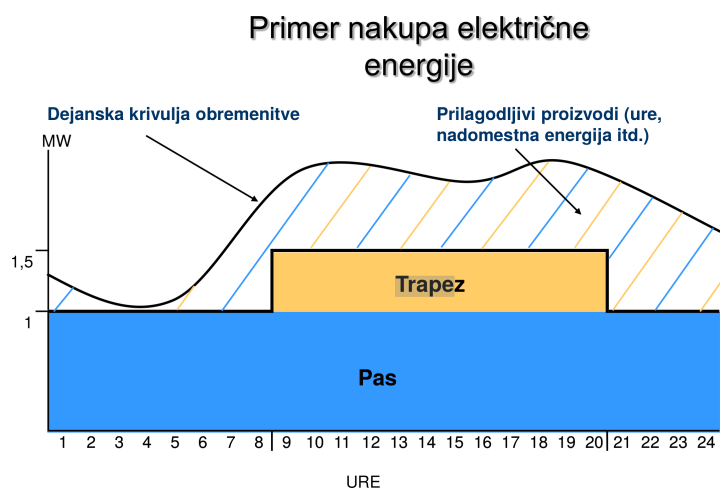
1.1 Cilji in motivacija

Z natančnejšim napovedovanjem porabe in proizvodnje električne energije, lahko zmanjšamo električne rezerve, ki so potrebna za glajenje nihanj v omrežju. Lahko zmanjšamo riziko preobremenitve sistema, kjer lahko pride do trajnih poškodb in okvar v omrežju. V primeru preobremenitve lahko odjemalci tožijo elektro distributerje za nastalo škodo, kar lahko za njih predstavlja velik finančen problem. Elektro distributerji lahko z dobro napovedano proizvodnjo prilagajajo svoja servisna dela tako, da z odklopom določenega območja ene ali več transformatorskih postaj, povzročijo čim manjšo izgubo lastnikom sončnih elektrarn. V prihodnosti bi lahko, v okviru dobrih napovedi in pametnih omrežij, bolj smiselno vklapljali in izklapljali električne naprave, kot so bojlerji, pralni stroji in podobno, glede na trenutno ceno elektrike.

Ambiciozni načrti glede nadaljnjega zmanjševanja emisij CO_2 nakazujejo, da se bo vključevanje obnovljivih virov še nadalje spodbujalo, poleg tega pa so določene tehnologije OVE že dosegle zrelost in postajajo cenovno vse bolj konkurenčne klasičnim tehnologijam. Deležniki, ki jih to zadeva, se bodo morali na spremenjene okoliščine prilagoditi ter najti poti za upravljanje vseh posledic, na primer preko povečane vloge pametnih omrežij in koncepta virtualnih elektrarn ter ne nazdanje boljšega napovedovanja proizvodnje.

1.1.1 Trgovanje na borzi z električno energijo pri nas in drugod

Vse večji delež proizvodnje iz obnovljivih virov energije ne prinaša velikih izzivov le na področju upravljanja elektroenergetskega sistema, temveč ima zelo izrazit vpliv tudi na dogajanje na elektroenergetskem trgu. Zaradi izjemno nizkih obratovalnih stroškov ter prednostnega dispečiranja, so se razmere na trgu z elektriko v zadnjih letih bistveno spremenile. To vpliva na tržne razmere, saj je prednostna uporaba tistih proizvodnih virov, ki obratujejo z najnižjimi obratovalnimi stroški ter glavnino prihodkov pridobivajo izven



Slika 1.2: Trapez kot borzni produkt [11].

trga. To so naprave iz obnovljivih virov energije, ki so cenovno neobčutljive, njihovo delovanje pa zasleduje le cilj čim višje proizvodnje glede na dane vremenske in ostale pogoje, ki vplivajo na obseg njihove proizvodnje.

Ključen vpliv porasta OVE je seveda čutiti na kratkoročnem trgu, torej na trgu za dan vnaprej in znotraj istega dne, kjer vremenske razmere in posledično proizvodnja iz obnovljivih virov energije predstavljajo temeljni dejavnik za ceno. Zaradi tega vpliva prihaja do velikih cenovnih razponov med posameznimi dnevi in celo urami, na določenih trgih pa prihaja tudi do pojava negativnih cen. Vpliv porasta deleža proizvodnje iz obnovljivih virov energije je čutiti tudi na dolgoročnem trgu [3].

Največji vpliv porasta OVE na tržne razmere na slovenskem trgu za dan vnaprej se kaže pri obratovanju sončnih elektrarn. Proizvodnja sončnih elektrarn je pri sončnem vremenu največja ravno takrat, ko so cene električne energije v preteklosti dosegale viške. To je v borznem produktu trapez (slika 1.2) na urnem nivoju med 8. in 20. uro. Kot rezultat tega se je dnevni diagram cen električne energije v zadnjih letih precej spremenil [3].

Del borznega trga predstavlja trgovanje na področju Slovenije, kar pomeni, da se trguje s produkti z dobavo v Sloveniji. Na tem segmentu trga

angleško	slovensko	čas
Base	Pas	00:00 - 24:00
Peak	Trapez	06:00 - 22:00
Off-peak1	Nočna 1	00:00 - 06:00
Off-peak2	Nočna 2	22:00 - 24:00
Euro-peak	Evro trapez	08:00 - 20:00
Euro-off-peak1	Evro nočna 1	00:00 - 08:00
Euro-off-peak2	Evro nočna 2	20:00 - 24:00

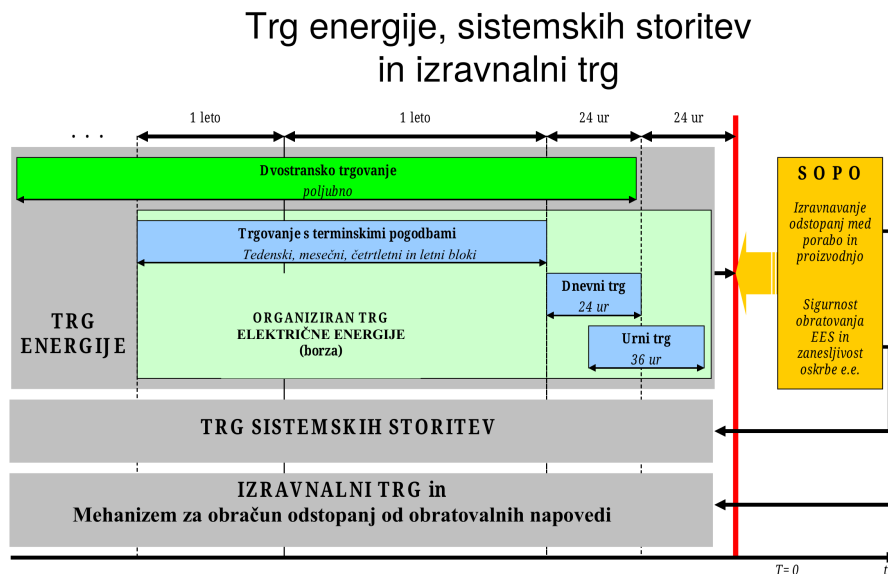
Tabela 1.2: Standardni produkti trgovanja na borzi.

lahko sodelujejo podjetja, ki so si to pravico pridobila z odobreno vlogo za sodelovanje na tem trgu [12]. Bilančna skupina je skupina, sestavljena iz poljubnega števila subjektov izravnave s pripadajočimi primopredajnimi mesti, ki se združijo zaradi skupnega obračuna odstopanj [11]. Trgovanje energije pomeni izmenjavo energije med bilančnimi skupinami. Odgovorni bilančne skupine pa odgovarja za odstopanja napovedi od realizacije.

Na borzi se trguje z standardiziranimi produkti naštetimi v tabeli 1.2. V sistemu trgovanja imamo različne trge (slika 1.3): trg energije, trg sistemskih storitev in izravnalni trg in mehanizem za obračun odstopanj od obratovalnih napovedi [11]. V trg energije je vključena borza s trgovanjem s terminskimi pogodbami, kjer lahko kupujemo elektriko v letnem, četrtnem, mesečnem, in tedenskem bloku. Vključen je tudi dnevni in urni trg. V realnem času pa se izvaja še izravnavanje odstopanj med porabo in proizvodnjo. Vse cene, tudi v letnem bloku, se določa za vsako uro posebej.

Trgovanje za dan v naprej

Trgovanje za dan na vnaprej na slovenskem borznem trgu poteka na način avckijskega trgovanja, pri čimer tržni udeleženci v fazi trgovanja v trgovalno aplikacijo EuroMarket vnašajo standardizirane urne produkte. Produkti so omejeni s cenovnim razponom od 0 €/MWh do 3000 €/MWh. Osnovna



Slika 1.3: Trgi elektroenergetskega sistema in njihova povezava. [11]

enota trgovanja (1 lot) na trgu za dan vnaprej znaša 1 MW/h [13]. Koliko stane napačna napoved je odvisno od borze. Avkcijsko trgovanje je razdeljeno na naslednje faze:

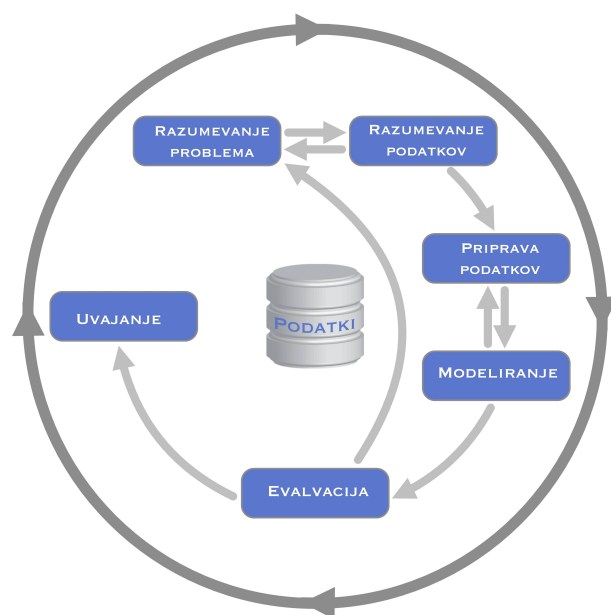
- faza trgovanja traja do 12:00 ure, ponudbe pa je možno vnašati že 8 dni pred samim dnevom trgovanja. V tej fazi je možen vnos, sprememba in preklic ponudb, udeleženci vidijo samo svoje ponudbe;
- faza mirovanja poteka od 12:00 ure do najkasneje 12:05 ure. V času faze mirovanja nadzornik trgovanja preveri vnešene ponudbe in ukrepa v primeru nepravilnosti;
- v fazi po trgovanju se izvede izračun marginalne cene, ki poteka med 12:05 uro in 12:52 uro, članom trgovanja se prikažejo marginalne cene, izračunane na avkciji;
- v neaktivni fazi je možen pregled marginalnih cen in lastnih poslov.

Čas, do kdaj je potrebno oddati napovedi, je odvisen od vsake bilančne skupine. Na primer podjetje E3, ki je vključeno v bilančno skupino Holding Slovenske elektrarne (HSE [14]), mora oddati dnevne napovedi do 8:00, podjetje GEN-i pa do 12:00. Napovedovanje je ključno na lokalni in globalni ravni. Na lokalni ravni je potrebno zadostiti določenim kvotam (SAIFI) in v primeru tehničnih okvar je potrebno poznati obremenjenost omrežja in možnih zančnih tokov. Na globalni ravni so potrebni letni zakupi prenosa elektrike med dvema lokacijama (STO) in v premru slabih napovedi imamo lahko tudi do 2x večji strošek od pričakovanega dobička. Ena izmed konkurenčnih prednosti med ponudniki električne energije kot so E3, GEN-i, Petrol, RWE, Elektro prodaja E.U. in drugih, je ravno v natančnejšem napovedovanju, saj lahko ponudijo nižjo, konkurenčnejšo ceno. Boljša napoved ne zmanjša le stroškov odstopanj, temveč koristi trgu kot celoti.

1.2 Metodologija

Za napovedovanje proizvodnje električne energije smo pridobili podatke o pretekli proizvodnji iz podjetja Elektro Primorska, o preteklih (lokalnih modelskih) vremenskih napovedih, pa iz agencije Republike Slovenije za okolje (ARSO). Napovedni modeli, ki smo jih razvili, so prilagojeni za območje Slovenske Primorske. Celotna metodologija podatkovnega rudarjenja je potekala po standardu CRISP-DM [15], ki je podrobneje razčlenjen na sliki 1.4.

Za izboljšanje natančnosti napovedovanja (oziroma zmanjšanje napake) smo uporabili različne metode podatkovnega rudarjenja: linearno regresijo z regularizacijo [16], regresijska drevesa, regresijske naključne gozdove [17] in K najbližjih sosedov (KNN). Napovedne modele smo gradili in testirali z orodjem scikit-learn [18]. Posebno pozornost smo namenili gradnji atributov (feature engineering) iz napovedanih prametrov meteorološkega modela ter izpeljanim atributom (zgodovinske in lokacijske). Preizkušali smo tudi razvrščanje (clustering) sončnih elektrarn v skupine glede na njihovo zgodo-



Slika 1.4: Diagram poteka metodologije CRISP-DM.

vinsko proizvodnjo električne energije.

1.3 Prispevek naloge

V nalogi smo predstavili problematiko in različne možne pristope pri obdelavi surovih izmerjenih podatkov. Preizkušali smo nekatere predlagane modele iz tujih člankov na naših podatkih in opazovali, kako se obnašajo s podatki elektrarn na primorski regiji Slovenije. Trenutno je v svetovnem merilu objavljenih 12 študij, ki pokrivajo dnevno napovedovanje proizvodnje fotonapetostnih modulov. Večino teh je bilo objavljenih v zadnjih 5 letih [19]. Razlog za pomanjkanje del na tem področju so težave s pridobitvijo podatkov operativnih fotovolotaičnih modulov zaradi varnostnih omejitev in pomanjkanja informacijske strukture za organizacijo in shranjevanje takšnih podatkov. Kot lahko vidimo v tabeli 2.1, smo uspeli za našo raziskavo pridobiti konkurenčno veliko podatkov na vseh področjih: številu elektrarn, skupni kapaciteti in časovnem intervalu [19]. V primerjavi z drugimi avtorji smo

poizkušali vključiti več modelskih točk, katerih vpliv smo modelirali glede na njihovo oddaljenost od sončne elektrarne. Iz vremenskih napovedi smo pridobili več atributov kot je običajno uporabljenih v člankih. Preizkušali smo napovedi z izgradnjo novih izpeljanih, zgodovinskih in lokacijskih atributov in rezultate ovrednotili tako v absolutnem merilu, kot v primerjavi s sorodnimi deli.

Poglavje 2

Pregled področja

Preden predstavimo raziskave pri nas in drugod, želimo poudariti, da je v literaturi možno zaslediti veliko načinov in modelov napovedovanja tako sončnega sevanja, kot končne izhodne moči fotovoltaičnih modulov. Ker gre za zelo korelirani vrednosti, smo v naši raziskavi uporabili ideje in modele iz obeh tipov raziskav. Zato smo tudi v tem poglavju predstavili in vključili oba tipa raziskav. Rezultate raziskav smo zaradi težkih medsebojnih primerjav predstavili opisno. Razlike med članki so v napovednem časovnem obdobju, metriki napak, dolžini napovedanega intervala, lokaciji elektrarn, napovedovanja sončnega sevanja ali pa proizvodnje sončnih modulov in različnih pogojih merjenja napake. Rezultate iz literature, ki so najbolj primerljivi z našo raziskavo, smo predstavili ter primerjali z našimi v poglavju 5.

2.1 Raziskave v Sloveniji

V Sloveniji se meritve sončne energije opravljajo le na nekaterih meteoroloških postajah. V centru odličnosti Vesolje-SI [20] so zato z lastnimi algoritmi določili prostorsko zvezno količino prejete sončne energije za vso Slovenijo, glede na fizikalne zakone ter detajlne podatke o reliefu in albedu. Tako so pridobili zemljevide osončenosti po karakterističnih mesecih. Izračunali so tudi optimalne orientacije sončnih celic za posamezne lokacije. Predlagajo

postavitve sončnih elektrarn in ugotavljajo njihovo donosnost. Izračuni so prilagojeni vsaki stavbi ter upoštevajo rastje in druge elemente, ki vplivajo na osončenost.

Na Fakulteti za elektrotehniko na univerzi v Mariboru (FERI) imajo lastno mikro sončno elektrarno in objavljeno spletno stran [21], kjer lahko v realnem času spremljamo podatke o proizvodnji ter meritve temperature in vetra. V okviru fakultete imajo razne diplomske naloge [22, 1] in doktorske disertacije. V enem izmed doktoratov [23] so poiskali trajektoriji naklona in azimuta, ki omogočata največjo proizvedeno električno energijo iz sončne elektrarne ob razpoložljivem sončnem sevanju. Pri izračunih in analizi razmer sledenja fotonapetostnega sistema soncu, potrebujemo čim bolj natančne vrednosti o razpoložljivi energiji sončnega sevanja v danem trenutku. V ta namen so izdelali tudi nov model za napoved sončnega sevanja. Da bi bil model čim bolj uporaben, so ga zapisali v obliki časovno odvisne funkcije. Sončno sevanje v obliki časovno odvisne funkcije so uporabili v optimizacijskem modelu za določitev optimalnih trajektorij sledilnega sistema. V sami optimizaciji so upoštevali tudi električne izgube oziroma porabo, ki nastanejo pri posameznih spremembah naklona in azimuta. Optimalno sledenje fotonapetostnega sistema soncu, ob upoštevanju izgub pogonskega sklopa, so določili z optimizacijskim algoritmom, imenovanim Diferenčna evolucija. Uporabe Diferenčne evolucije za iskanje optimalnih trajektorij sledilnega sistema takrat v literaturi še ni bilo mogoče zaslediti, tako da je bil v tem delu za ta namen uporabljen prvič [23].

Tudi v Ljubljani v okviru Fakultete za elektrotehniko (FE) deluje Laboratorij za fotovoltaike in optoelektroniko, ki se ukvarja z dolgoročnim napovedovanjem energijskega izplena sončnih elektrarn. Bodočim investorjem želijo prikazati, koliko lahko pričakujejo od svoje investicije, da lahko na podlagi tega tudi ocenijo njeno energijsko in predvsem ekonomsko smiselno. V primeru napovedi energijskega izplena PV modulov ali sistema le na osnovi STC ali NOCT vrednosti, podanih v podatkovnih polah proizvajalcev, namreč pogosto prihaja do preveč optimističnih ocen izplena [24]. Optimistična

napoved sicer pomaga pri odločitvi o investiciji, vendar to dolgoročno pomeni razočaranje, ker ta ne dosega prej napačno (preveč optimistično) predvidenega energijskega izplena [25]. V okviru FE je bilo letos objavljeno delo, podobno našemu, na področju kratkoročnega napovedovanja proizvodnje električne energije sončnih elektrarn in malih hidroelektrarn, kjer so skušali z nevronskimi mrežami napovedovati 24-urno proizvodnjo. V njihovih modelih so napovedovali proizvodnjo gorenjske regije za 25 sončnih elektrarn [26].

2.2 Tuje raziskave

V enem izmed člankov [19], ki so bili objavljeni leta 2016 na področju dnevnega napovedovanja električne proizvodnje iz fotonapetostnih modulov, so raziskovalci predstavili problematiko premajhnega raziskovanja področja. V zadnjih 5 letih je bilo na to temo objavljenih 12 študij, ki pokrivajo napovedovanje proizvodnje za dan vnaprej iz fotonapetostnih modulov. Razlog za primanjkljaj so težave s pridobitvijo podatkov operativnih fotonapetostnih modulov zaradi varnostnih omejitev in pomanjkanja informacijske strukture za organizacijo in shranjevanje takšnih podatkov. V tabeli 2.1 je prikazana primerjava podatkov različnih člankov, s katero dobimo lažjo predstavbo za kakšne velikosti raziskav sploh gre. V tem članku so kot doprinos vrednotili medletne napake napovedi in predstavili prostorsko glajenje (spatial smoothing) s pari najbližjih elektrarn.

Ker so v večini tujih člankov kot dosedanje delo omenjeni različni statistični modeli časovnih vrst, vključno z auto-regresivnimi (AR) modeli, tekočim povprečenjem (MA) in njihovimi generaliziranimi verzijami ARMA in ARIMA modeli, bomo predstavili članek [38], kjer je predstavljen tudi model ARMAX. V primerjavi z drugimi napovednimi tehnikami, model časovne vrste ARIMA ne potrebuje meteoroloških napovedi. Zaradi te preprostosti, je bil statističen model ARIMA najširše predstavljen in uporabljen. Intuitivno so take informacije zelo pomembne pri izboljšanju napovedne točnosti, zato so v tem članku predlagali generaliziran model ARMAX, ki dovoli zu-

Regija	število elektrarn	skupna kapaciteta	dolžina podatkov
Danska [27]	21	100 kWp	1 leto
Španija [28]	1	6 kWp	10 mesecev
Nemčija [29]	11	/	2 meseca
Nemčija [30]	383	100 MWp	10 mesecev
Japonska [31]	454	10 MWp	1 leto
Japonska [32]	273	10 MWp	1 leto
Kanada [33]	3	100 kWp	2 leti
Francija [34]	28	/	2 leti
Francija [35]	28	/	2 leti
Španija [36]	5	10 MWp	2 leti
Italija [37]	1	20 kWp	4 dni
Kitajska [38]	1	2 kWp	6 mesecev
Primorska	211	15 MWp	12 mesecev

Tabela 2.1: Primerjava podatkov iz literature.

nanje vhode v napovedni model. Predlagani model potrebuje temperaturo, količino padavin, čas osončenja in vlažnost, katere je enostavno pridobiti iz lokalnih vremenskih postaj. Ker se model ARMAX ne zanaša na sončno sevanje, ohranja preprostost kot model ARIMA. Po drugi strani je bolj splošen in fleksibilen za praktično uporabo. Model ARMAX zelo izboljša napovedi v primerjavi z modelom ARIMA. V članku navajajo tudi, da je njihov model boljši od nevronske mreže.

V članku [39] so predstavili primerjavo fiksnih modelov, ki so naučeni samo enkrat, in premikajočih modelov (moving models), kjer je učna množica n zadnjih oziroma n podobnih dni. V rezultatih so pokazali, da so z ključnimi gozdovi imeli manjšo kvadratno absolutno napako od regresijskih podpornih vektorjev, model KNN in linearno regresijo pa so uporabili kot njihov osnovni (baseline) model. Pokazali so tudi, da so premikajoči modeli bolje napovedovali sončno sevanje od fiksnih modelov.

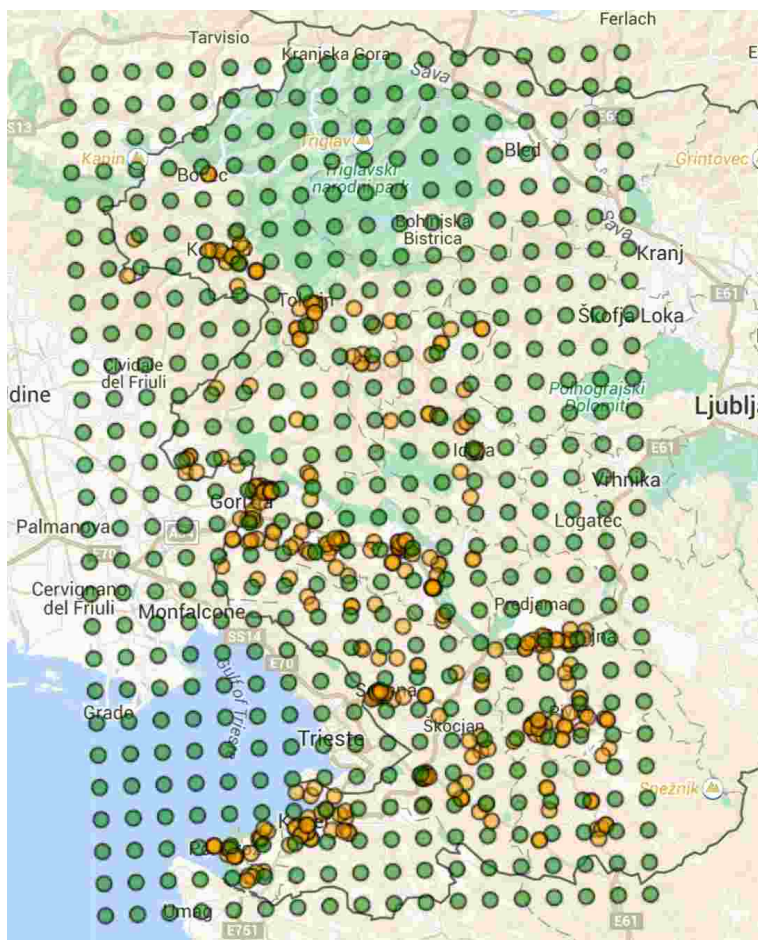
V članku [40] so predstavili napovedovanje sončnega sevanja z dvema nevronske mrežama, ločenima za globalno in direktno sončno sevanje. V predlaganem modelu niso uporabljali zgodovinskih podatkov. Rezultate so primerjali le z ostalimi članki, ki so uporabljali nevronske mreže.

Poglavje 3

Podatki

V okviru prvih dveh korakov po metodologiji CRISP-DM, razumevanje problematike in razumevanja podatkov, smo opravili precej sestankov na podjetjih Elektro Primorska, E3 in ARSO. Uporaba metodologije CRISP-DM nam je omogočala lažjo predstavitev postopka na omenjenih podjetjih. Potrebni so bili razni dogovori in utemeljevanja, da smo podatke lahko pridobili. Utemeljiti smo morali namen uporabe podatkov in predstaviti, kako bi naša raziskava pripomogla k izboljšanju poslovanja podjetij. Skozi celoten proces podatkovnega rudarjenja, smo bili s podjetji v stalni komunikaciji preko elektronske pošte, saj so se različne nejasnosti in napake pokazale šele ob filtriranju in analizi podatkov ter ob samem modeliranju. Vsi podatki skupaj, ki smo jih prejeli od podjetij in shranili v podatkovno bazo, obsegajo 40.731.416 zapisov.

Podatki pokrivajo celotno območje primorske regije v Sloveniji, kar je vidno na sliki 3.1. Z zeleno so označene lokacije Aladinovih modelskih točk, z oranžno pa lokacije sončnih elektrarn. Zaradi pomanjkanja zgodovinskih modelskih vremenskih napovedi smo modele gradili samo za obdobje dveh let (2011, 2012).



Slika 3.1: Prikaz modelskih točk z zeleno in sončnih elektrarn z oranžno barvo na zemljevidu.

3.1 Proizvodnja elektrarn

Za pridobitev podatkov o zgodovini proizvodnje sončnih elektrarn smo morali najprej pripraviti izjavo o varovanju podatkov. Zelo pomembno je bilo, da smo dobili čim več podatkov, saj so za podatkovno rudarjenje ključni. Elektrarne imajo različne datume priklopa in nekatere so bile priključene pozneje, zato je količina podatkov za posamezno elektrarno različna.

Podatki, ki smo jih prejeli, zavzemajo 2.7 GB prostora. Obsegajo 345 elektrarn od leta 2011 do vključno leta 2015. Za vsako elektrarno smo dobili: številko merilnega mesta, datum priklopa, naslov proizvodne naprave, velikostni razred proizvodne naprave, količino letno proizvedene energije, in nazivno moč proizvodne naprave (priključna moč). Meritve so shranjene na vsakih 15 minut, zato imamo 96 meritev na dan. Meritve so v kilovatih ali pa v kilovatnih urah, odvisno od starosti števca. Števci, katerih meritve so v kilovatnih urah, strogo naraščajo (akumulirajo). Pomembno je, da se meritve ne pošiljajo v realnem času, ampak se ponavadi pošljejo vsako uro (4 meritve) ali pa enkrat na dan, odvisno od števca in velikosti elektrarne. Ugotovili smo, da so lahko nekatere meritve neustrezne, zato smo jih na podlagi analize, opisane v razdelku 3.1.1, popravili ali odstranili. Naš končni kriteriji za določanje napake smo določili tako, da je napačna meritev tista, ki nima oznake statusa 0 (razdelek 3.2), ki je 3x večja od priključne moči in ko je v dnevu že 15% nevalidiranih podatkov. Izluščili smo 211 elektrarn, ki smo jih klasificirali kot primerne za nadaljnjo podatkovno rudarjenje.

3.1.1 Prostorsko označevanje podatkov

Točnih koordinat za vse elektrarne podjetje Elektro Primorska nima. Za nekatere elektrarne smo dobili koordinate v geografskih kartezičnih koordinatah Gauss-Krügerjevega sistema. Zaradi konsistentnosti s koordinatami modela ALADIN, jih je bilo potrebno pretvoriti v sistem WGS84. Gauss-Krügerjev koordinatni sistem je stari uradni državni koordinatni sistem v Sloveniji. Transverzalni Mercatorjev koordinatni sistem je položen preko Bessel-ovega

lastnost	vrednost
število elektrarn	345
časovni obseg podatkov	1.1.2011 - 16.11.2015
časovni interval meritve	15 minut
število meritev na dan	96
posodobitev podatkov v bazi	vsako uro (4 meritve) ali enkrat na dan odvisno od števca
osnovna enota meritve	kWh in kW
velikost na disku	2.7 GB
skupna moč	35.91 [MW]
povprečna moč	104 [kW]
atributi	številka merilnega mesta, datum priklopa, naslov proizvodne naprave, velikostni razred proizvodne naprave, količina letno proizvedene energije, nazivno moč proizvodne naprave (priključna moč)

Tabela 3.1: Lastnosti prejetih podatkov sončnih elektrarn. (vir: Elektro Primorska)

lokalnega elipsoida. Njegovi parametri so bili določeni tako, da se kar najbolje prilega področju nekdanje Jugoslavije. Sistem WGS84 pa je referenčni elipsoid, določen leta 1984. Določen je tako, da se najbolje prilega Zemlji v celoti, torej da je vsota odmikov površja Zemlje od tega elipsoida minimalna. Uporablja se ga v povezavi z GPS sistemom satelitske navigacije [41]. Za pretvarjanje smo uporabili knjižnico Geo Coordinate Converter [42], ki omogoča pretvarjanje s Helmertovo transformacijo [43]. Za nekatere elektrarne smo pridobili le koordinate najbližje transformatorske postaje (TP), kar lahko povzroči napako do 300 metrov. Ostale manjkajoče podatke smo pridobili s pomočjo naslova in hišne številke. Pomagali smo si z javno dostopnimi servisi zemljevidov Bing, zemljevidov Google in servisom Nominatim, ki je dostopen v okviru projekta OpenStreetMap [44]. S pomočjo teh servisov je iz naslova možno pridobiti koordinate v tekstovni obliki.

3.1.2 Spoznavanje in čiščenje podatkov

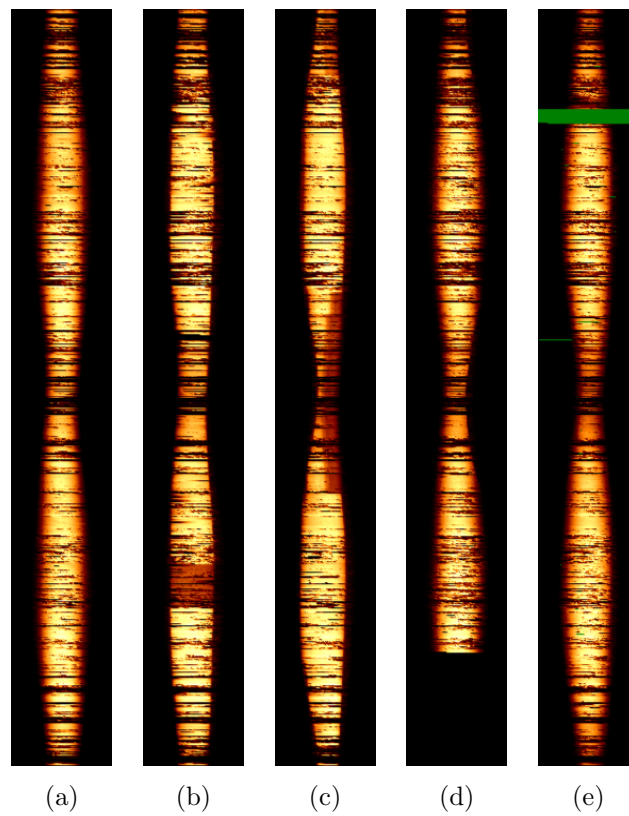
Ko smo podatke pridobili, nikakor nismo mogli vedeti če so sploh smiselni. Iz članka [45] smo dobili idejo, da predstavimo podatke z vretenasto sliko, kjer horizontalna (x) os predstavlja minute in vertikalna (y) os predstavlja dni. Svetlost piksla predstavlja normalizirano proizvodnjo elektrarne. Tako enostavno interpretiramo dnevno in sezonsko obnašanje sončnega sevanja. Črna barva pomeni, da elektrarna ni proizvedla nič elektrike. Prehod med črno in rumeno pa se kaže v obliki rdečkaste barve in je pokazatelj, da se sončno sevanje, ki pada na površino celic, večja ali manjša. Z zeleno barvo smo označili napačne meritve. V članku [45] je predstavljen sistem napovedovanja sončnega sevanja na podlagi takšnih slik.

Za vsako elektrarno smo torej izrisali vretenasti diagram in ga ročno pregledali. Opazili smo veliko anomalij. Na sliki 3.2 je prikazanih pet izrisov elektrarn za obdobje dveh let. Glede na prvo vreteno (a) lahko sklepamo, da gre za tipično elektrarno s pričakovano proizvodnjo. Na vretenu (b) vidimo, da je bila v prvem letu med poletjem proizvodnja nižja od pričakovane. S posvetovanjem z lastniki sončnih elektrarn smo ugotovili, da je to možno

zaradi delne poškodbe elektrarne, torej je nekaj celic v okvari, lahko pa so bile celice zelo umazane in je zato proizvodnja padla. Če gre za večjo elektrarno, se lahko tudi pokvari kateri izmed razsmernikov in je tako znižana maksimalna proizvodnja celotne elektrarne. V primeru okvare sončnih celic oziroma razsmernikov lastniki unovčijo garancijo, vendar lahko traja tudi do dva meseca, preden se opravi celotno popravilo. Pri enem izmed lastnikov se je celo zgodilo, da so podgane požrle dovod do števca in je bil lastnik zato nekaj dni brez dohodka. Vsi takšni dogodki se lahko zelo lepo vidijo na vretenastih izrisih. Na vretenu (c) opazimo znižano proizvodnjo v zimskem času le popoldne. Sklepamo lahko, da gre za slabo postavljene fotonapetostne module. Večino lastnikov elektrarn jih postavi kar na streho, kjer ni nujno najbolj ugoden naklon za proizvodnjo. Ko sončni žarki ne padajo več pod optimalnim kotom, proizvodnja drastično pade. Na vretenu (d) se jasno vidi, da je bila elektrarna postavljena pozneje od ostalih in je zato njena proizvodnja v začetnem obdobju nič. Na vretenu (e) smo z zeleno barvo označili napake, kot bomo podrobneje videli v nadaljevanju. Pri nekaterih elektrarnah smo zaznali napačno sinhronizacijo ure, saj je bila proizvodnja prikazana ponoči, čez dan pa ne. Nekatere elektrarne so imele podatke v akumulacijski obliki in njihova vrednost proizvodnje ni predstavljala trenutne, ampak akumulacijo proizvodnje od začetka postavitve elektrarne čez vsa leta. Take elektrarne smo ročno označili in podatke pretvorili v neakumulacijsko obliko.

Obravnavanje napak

Dokument Standardizirani merilni in obračunski podatki [46] določa enoten standardizirani zapis podatkov in način posredovanja merilnih in obračunskih podatkov s strani izvajalcev nalog Systemskega operaterja distribucijskega omrežja (SODO) na celotnem področju Republike Slovenije. Tudi Elektro Primorska ima podatke shranjene po tem standardu. Pri vsaki meritvi je zabeležen status meritve, ki ima lahko eno izmed devetih vrednosti, ki smo jih zapisali v tabeli 3.2. Prešteli smo 12.645.826 meritev, ki so bile napačne (enačba (3.1)). Validirane in nevalidirane meritve so meritve, ki jih števec na



Slika 3.2: Vretenast prikaz proizvodnje za pet različnih elektrarn v obdobju dveh let. Z belo barvo je označena najvišja proizvodnja, s črno pa najnižja. Z zeleno barvo so označene napačne meritve. Širina belooranžne barve predstavlja dolžino sončnega sevanja na elektrarno.

<i>status</i>	<i>opis</i>
0	validiran
1	nevalidiran
2	nadomesten podatek
3	popravljen podatek
4	simulirani podatki
5	agregirani podatki
6	podatek manjka
7	napaka po validaciji
8	error

Tabela 3.2: Standardne oznake izmerjenih podatkov proizvodnje.

merilnem mestu označi kot take z uporabo internih validacijskih sistemov. V zadnjem koraku informacijski sistem omogoča ekspertu naknadno preverjanje in validacijo prek grafičnega uporabniškega vmesnika, da ne prihaja do napačnih zneskov ob obračunu. Žal smo ugotovili, da ročna validacija običajno ni bila uporabljena in smo jo zato skušali izvesti sami.

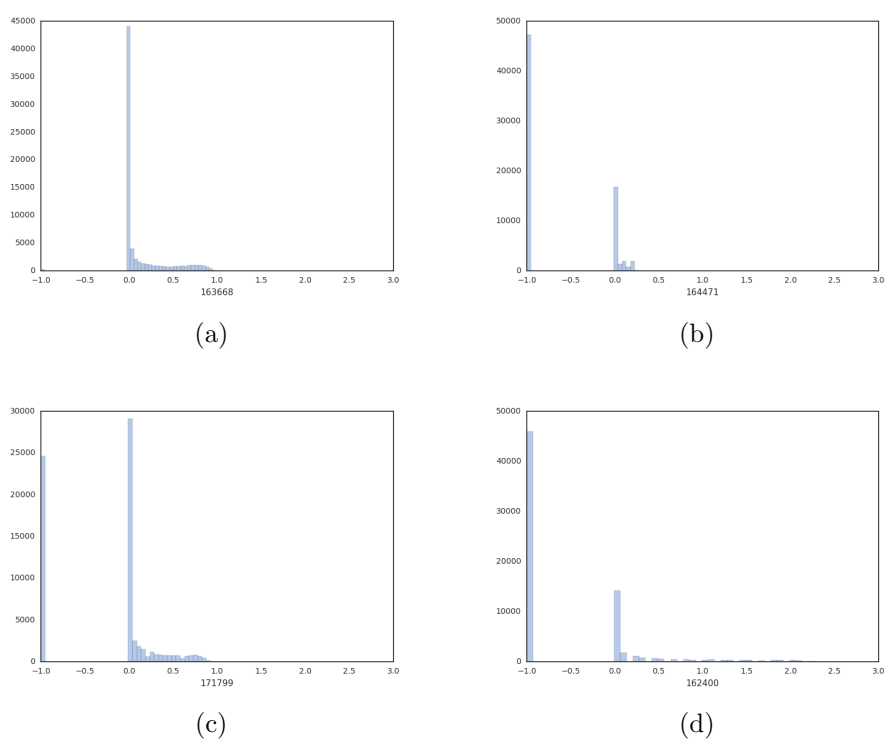
$$Napaka = status \neq 0 \quad (3.1)$$

Zaradi presenetljivo velikega števila napak, smo začeli podatke analizirati z različnimi histogrami. Za vsako elektrarno smo naredili tudi histogram relativne (glede na nazivno moč) proizvodnje. Na napačnih meritvah smo označili relativno proizvodnjo z -1. Opazili smo, da je lahko tudi proizvodnja večja od 1, zato smo se pozanimali, na kakšen način je določena priključna moč. Lastnik elektrarne mora za pridobitev soglasja podati priključno moč elektrarne. To pomeni, da ima lahko elektrarna manjšo proizvodnjo od priključne moči in navedene naj ne bi presegla. Ugotovili smo, da imajo nekatere elektrarne tudi do trikrat večjo proizvodnjo od maksimalne, saj lastniki lahko malo prirejajo podatke. Na podlagi 3D histograma proizvodnje vseh elektrarn (slika 3.4) in 2D histograma proizvodnje vseh elektrarn (slika 3.5)

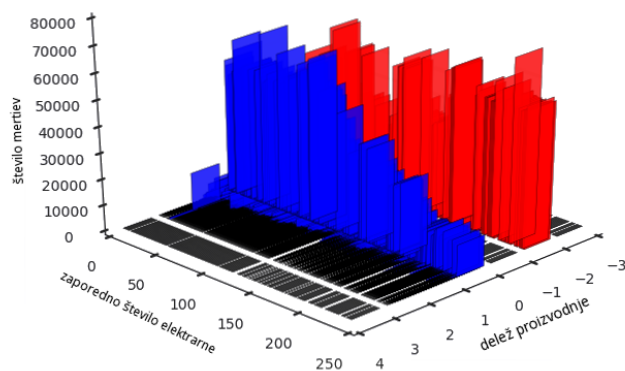
smo se odločili, da izračunamo novo priključno moč glede na največjo proizvodnjo elektrarne v celotnem obdobju, da so porazdelitve proizvodnje elektrarn lahko bolj primerljive. Nekaj meritev je bilo tudi s 100-krat večjo proizvodnjo od maksimalne. Take meritve smo označili kot napačne. Na sliki 3.3 vidimo primer štirih elektrarn, ki imajo normalizirano proizvodnjo pred popravljeno maksimalno proizvodnjo. Na navpični osi je število meritev za določen stolpič histograma. Na vodoravni osi pri vrednosti -1 so vse nevalidirane meritve. Vidimo, da je lahko tudi polovico meritev označenih kot nevalidiranih. Za lažjo predstavo smo vse takšne histograme združili na en 3D histogram (slika 3.4). Nevalidirane meritve so označene z rdečo barvo ostale pa z modro. Pri glavnem dobavitelju števecov iz podjetja Landis smo povprašali, kako je možno, da prihaja do tako prevelikih vrednosti meritev, ki jih števec vseeno označi kot veljavne. Glavni razlog za veliko špico v določeni meritvi je običajno ob montaži oziroma menjavi števca. To se zgodi zaradi tokovnih prestav. Tokovne prestave omogočajo natančnejše meritve, vendar ob montaži lahko pride do velikih električnih nihanj, tudi do takšnih za katere tokovne prestave niso namenjene. Nekaj neobičajnih meritev, ki smo jih ročno preverili pri dobavitelju, je nastalo na števcu *MT830* proizvajalca *Iskraemeco*.

Nato smo začeli preučevati napake in poiskali možnosti za zmanjševanje njihovega vpliva. Napake lahko označi merilna naprava ob sami meritvi, tiste, ki ne označi kot napako, pa moramo zaznati sami, saj s tem zmanjšamo napako kasneje pri napovedi. Napake smo izrisali z zeleno barvo, kot je razvidno iz vretena (e) na sliki 3.2.

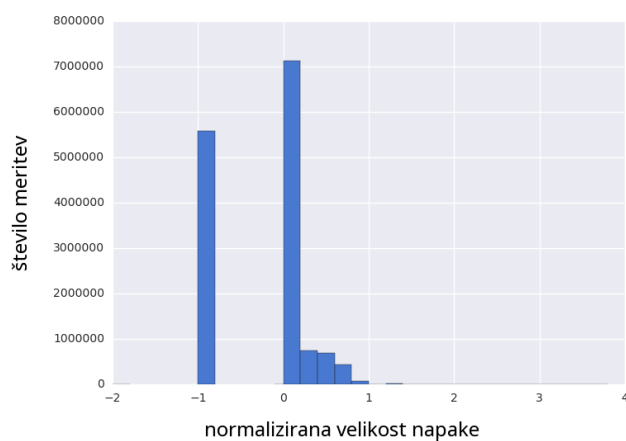
Vpliv napak smo skušali zmanjšati tako, da smo napačne podatke odstranili ali pa smo jih generirali. Napačni podatki so podatki, ki so že bili označeni kot napačni (enačba (3.1)) skupaj s tistimi, ki smo jih mi označili kot napačne. V ta namen smo izrisali histogram napak na dan za vsako elektrarno (slika 3.6), kjer vertikalna os (y) predstavlja število dni, horizontalna os (x) pa število napak v dnevu. Naredili smo tudi histograme, kjer je razvidno število napak za vsak dan posebej (slika 3.7), kjer so na horizontalni



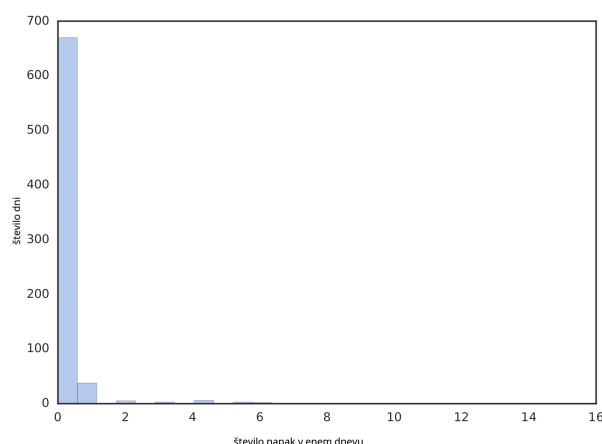
Slika 3.3: Histogrami normalizirane proizvodnje štirih elektrarn z napakami na vrednosti -1.



Slika 3.4: Normalizirana proizvodnja elektrarn na enem 3D histogramu, kjer višina stolpca predstavlja število meritev, širina pa normalizirano proizvodnjo ene elektrarne. Z rdečo barvo so predstavljene napačne meritve, z modro pa ostale.



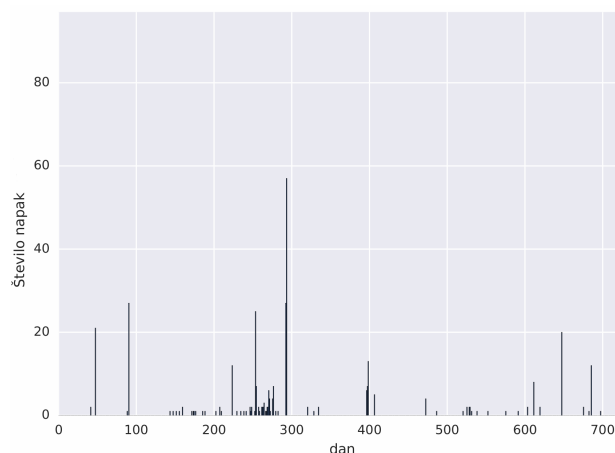
Slika 3.5: Skupna normalizirana proizvodnja vseh elektrarn na enem histogramu. Višina stolpca predstavlja število meritev, lokacija (pozicija) pa normalizirano proizvodnjo. Stolpec, ki ima na horizontalni osi vrednost -1, predstavlja napačne meritve.



Slika 3.6: Primer histograma napak na dan za eno elektrarno, kjer je bilo največ 16 napak v dnevu.

osi (x) dnevi v dveh letih in na vertikalni osi število napak. Iz slednjih smo opazili, da imajo nekatere elektrarne napake le na začetku meritev. Ugotovili smo, da to velja za tiste lastnike, ki so podpisali pogodbo, elektrarne pa so postavili kasneje. V takih primerih smo zmanjšali število napak tako, da smo vsem elektrarnam prestavili začetni datum priklopa na prvi dan veljavne proizvodnje. Na tak način smo odstranili veliko število napačnih meritev.

Poizkušali smo združiti posamezne histograme za vsako elektrarno, da bi lahko dobili boljši pregled nad celoto. Na histogramu 3.8 je viden delež napak na dan vseh elektrarn skupaj. Ustvarili smo ga tako, da smo naredili histogram napak za vsako elektrarno posebej in ga normalizirali tako, da smo ga delili s številom vseh napak te elektrarne. Na koncu smo vse normalizirane histograme sešteli in ponovno normalizirali z vsoto vseh stolpcev zato, da je vizualizacija lažje razumljiva. Na podlagi tega histograma smo se odločili, da v primeru večjega števila napačnih meritev nad 15 v enem dnevu po interpolaciji, ta dan označimo kot “slab” dan in ga odstranimo. V članku [39] so se sicer odločili za mejo 25% meritev, kar bi v našem primeru znašalo 24 meritev, vendar je njihova izbira parametra neutemeljena, torej izbire ne moremo enostavno primerjati. Z izločanjem slabih dni se lahko izognemo

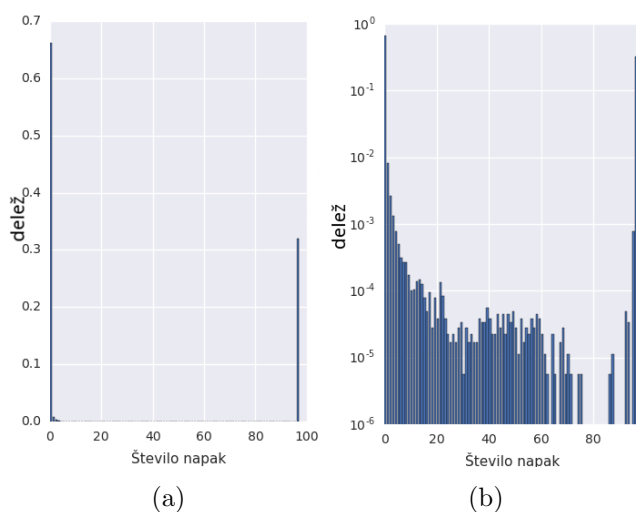


Slika 3.7: Dvoletni histogram števila napak na dan.

problemu neuspešne interpolacije. Ko smo končali z označevanjem napak, smo namreč napačne podatke nadomestili z rezultati linearne interpolacije. Pri tem smo uporabljali orodje *scipy* [47].

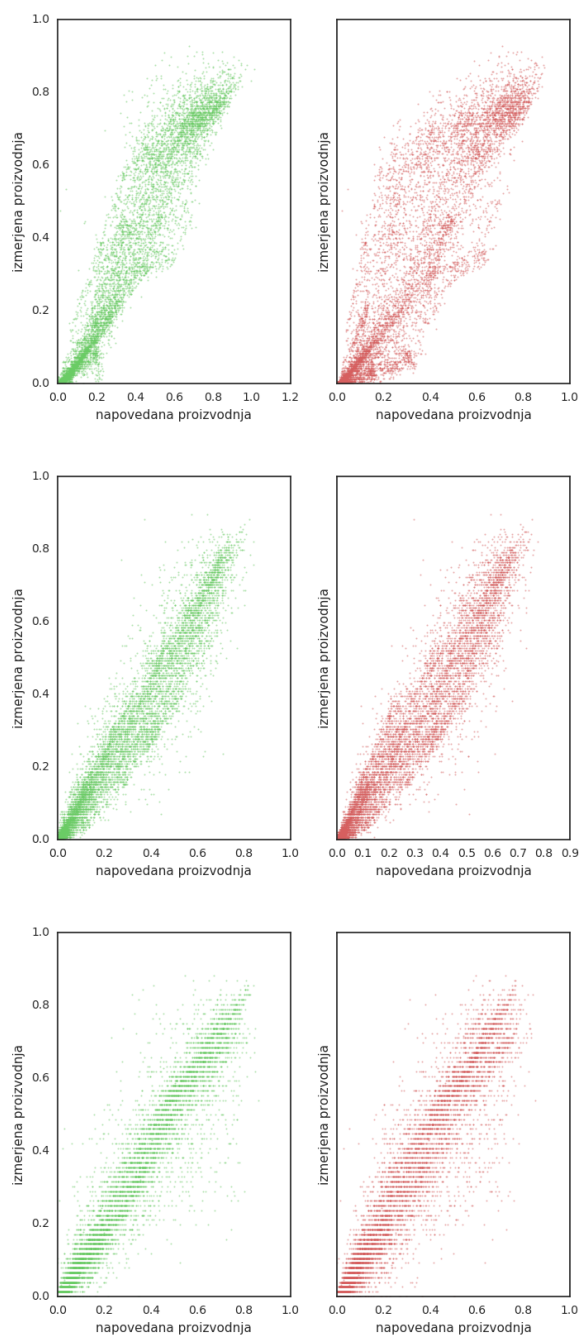
Zaradi zgoraj opisanih možnih vzrokov napak, smo skušali tudi take napake ugotoviti s pomočjo proizvodnje sosednjih elektrarn. Za vsako elektrarno smo poiskali pet lokacijsko najbližjih. Ker je vreme na krajših razdaljah zelo podobno, smo sklepali, da je podobna tudi relativna proizvodnja. V ta namen smo zgradili dva preprosta modela, model povprečja in model linearne regresije. Model smo na sosednjih elektrarnah naučili in napoved primerjali z opazovano elektrarno. V času, ko je imela opazovana elektrarna nevalidirane vrednosti, smo proizvodnjo iz učnih in napovednih podatkov odstranili. V primeru validirane vrednosti na opazovani elektrarni in nevalidirane vrednosti na sosednji elektrarni, smo podatek zamenjali z lokalno napovedjo (linearno ali povprečje) sosedov. S pomočjo *grafa raztrosa* (scatter-plot) (slika 3.9) smo poizkušali določiti smiselno mejo, koliko mora napoved odstopati, da meritev označimo kot napako. Ker ni videti jasne meje, smo se odločili, da to ne bo mogoče.

Žal smo bili zaradi časovnega obsega vremenskih podatkov (razdelek 3.2) omejeni na leti 2011 in 2012. Na podjetju Elektro Primroska so nam povedali, da naj bi se v zadnjih letih število nevalidiranih meritev zmanjšalo, saj

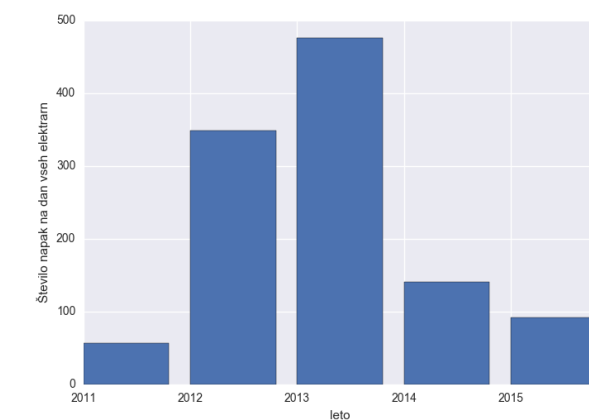


Slika 3.8: Delež napak na dan vseh elektrarn skupaj na linearni (a) in logaritemski skali (b). Do ekstremnih vrednosti prihaja zaradi daljšega nedelovanja elektrarn.

se je veliko števcov moderniziralo in se je informacijski sistem shranjevanja izboljšal. Pregledali smo število napak, za vse podatke do leta 2015. Na histogramu 3.10 smo predstavili število napak za 211 elektrarn, ki smo jih uporabljali pri modeliranju. Število napak na dan smo dobili tako, da smo prešteli vse napačne meritve vseh elektrarn in jih delili s številom dni v letu. Razlog, da je v letu 2011 in 2012 napak nenavadno manj kot v letu 2013, je razlog ta, da se je veliko elektrarn priključilo šele v drugi polovici leta 2012.



Slika 3.9: Trije grafi raztrosa, kjer je na eni osi izmerjena vrednost, na drugi pa vrednost napovedana z modelom povprečenja (na desni) in linearno regresijo (na levi). Prikazujejo odvisnost napak napovedi, ustvarjenih na podlagi najbližjih sosedov, od izmerjene vrednosti.



Slika 3.10: Povprečno število napak na dan za vse elektrarne od leta od 2011 do 2015. Povečanje napak do leta 2013 je posledica sprotnega priklapljanja elektrarn.

3.1.3 Organizacija podatkov

Za lažje delo, organizacijo in validacijo podatkov smo podatke shranili v podatkovno bazo MySQL. Tak način nam je omogočal lažji pregled nad spremembami v podatkih. Ustvarili smo pet shem. Vsaka shema predstavlja eno stopnjo (migracijo).

1. surove podatke smo samo skopirali iz datotek tipa .csv, .xlsx in .xls, take kot so, brez modifikacij;
2. identificirali in pretvorili smo akumulacijske podatke elektrarn in vremenske podatke sončnega sevanja. Nastavili smo tudi začetni datum proizvodnje po zgoraj opisanem postopku;
3. označili smo napačne dni, kjer pa je bilo mogoče, smo nove podatke izračunali s pomočjo linearne interpolacije;
4. odstranili smo elektrarne, kjer je veljavnih dni manj kot 150, 15-minutno proizvodnjo pa smo povprečili, da smo dobili enourno proizvodnjo;
5. za vsako elektrarno smo modelski napovedi poiskali podobno modelsko napoved iz preteklosti (razdelek 4.1).

3.2 Vremenske napovedi

Podatke vremenskih napovedi za leti 2011 in 2012 smo pridobili od Agencije ARSO [48]. Žal modelskih napovedi nismo mogli pridobiti za ostala leta do 2015, ker ponovni izračunani še niso bili končani. Agencija ARSO namreč ne hrani vseh zgodovinskih podatkov napovedi zaradi pomanjkanja diskovnega prostora, hrani pa vhodne parametre za izračun modela. Na srečo so letos testirali novo verzijo modela ALADIN s preteklimi vhodnimi podatki (tudi za leti 2011 in 2012), tako da smo lahko dobili izhodne izračune modela, ki smo jih potrebovali.

3.2.1 Opis podatkov

Podatke smo dobili v 21.930 datotekah .csv. Vsaka datoteka predstavlja eno uro v letih 2011 in 2012. V imenu datoteke smo dobili časovno značko, v kateri so napovedi vseh modelskih točk. Napovedni parametri so prikazani v tabeli 3.3.

Napovedi se generirajo enkrat na dan za 30 ur v naprej. Zavzemajo 450 MB prostora in obsegajo 486 modelskih točk. Napovedi za vsako točko so izračunane s pomočjo numeričnega mezo-meteorološkega modela ALADIN/SI [49]. Model ima resolucijo 4.4 km, kar pomeni, da vrednost v modelski točki predstavlja povpečno vrednost za površino 19.36 km^2 (slika 3.11). Nekaterne vremenske parametre smo morali pretvoriti v primernejšo obliko:

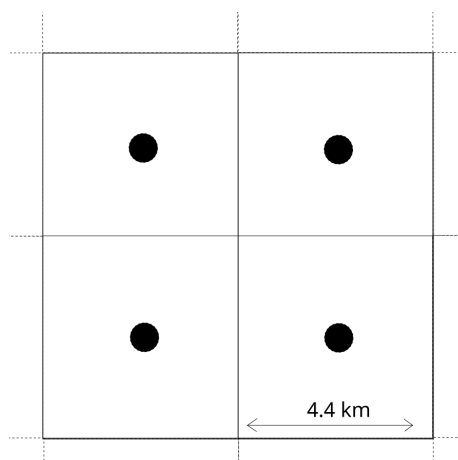
- povprečno jakost sevanja (W/m^2) v enournem intervalu dobimo tako, da odštejemo dve zaporedni vrednosti kumulativnega sevanja in rezultat delimo s časom (3600 sekund);
- geopotencial je potencial zemeljskega gravitacijskega polja. Lahko si ga predstavljamo kot razliko med potencialno energijo nekega objekta na določeni nadmorski višini in potencialno energijo objekta z enako maso na morski gladini. Če ga delimo z ~ 9.8 dobimo nadmorsko višino modelske točke v metrih. Ker je opis reliefa v modelu opisan spektralno, lahko pride tudi do negativnih vrednosti [50];

3.2.2 Ugotovitve

Akumulacijske napovedi smo morali pretvoriti v trenutne vrednosti. Pri pretvarjanju smo ugotovili, da se vrednosti lahko občasno tudi zmanjšajo, vendar napovedano negativno sevanje ni smiselno. Ugotovili smo, da problem nastane le v določeni uri pri sončnem zahodu, ko nad določenim delom celotnega modelskega izračuna sonce še sije, nad ostalim pa ne sije več 3.12. Negativne razlike se ne pojavijo pri neto sončnem sevanju pri tleh, ki je osnovna količina pri shemi za sevanje. Izvor je v naravi numeričnega izračuna. Ne pojavlja

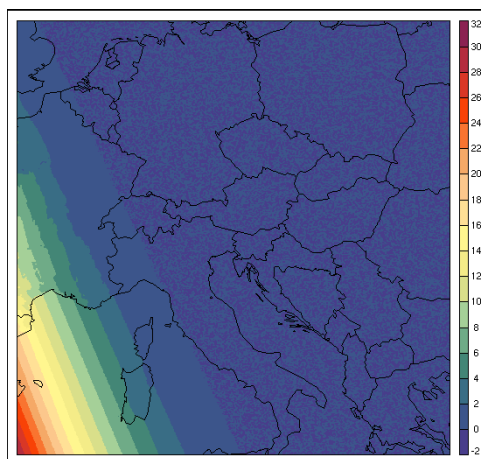
napovedni atributi vremena	enota mere
zemljepisna širina	° (stopinja)
zemljepisna dolžina	° (stopinja)
temperatura zraka na dveh metrih	K
relativna zračna vlaga na dveh metrih	%
oblačnost ali delež pokritosti	%
u komponenta vetra	m/s
v komponenta vetra	m/s
kumulativna količina padavin od izračuna napovedi do določene ure napovedi	l/m^2
kumulativno globalno kratkovalovno sevanje pri tleh	J/m^2
kumulativno neto kratkovalovno sevanje pri tleh	J/m^2
kumulativno neto kratkovalovno sevanje na vrhu atmosfere	J/m^2
kumulativno neto dolgovalovno sevanje pri tleh	J/m^2
kumulativno globalno dolgovalovno sevanje pri tleh	J/m^2
geopotencial	J/kg
kopno / morje	1 / 0

Tabela 3.3: Atributi modelskih točk



Slika 3.11: Štiri modelske točke na ravnini.

se zjutraj pri vzhodu, ampak le občasno pri sončnem zahodu v shemi globalnega sevanja. Ker so negativne vrednosti zelo majhne, se lahko postavijo na nič. Največja urna gostota energijskega toka v času negativnih vrednosti je namreč pod 500 J/m^2 , kar je $\sim 0.14 \text{ W/m}^2$.



Slika 3.12: Širši pregled zemljevida osvetljenosti ob sončnem zahodu. (vir: ARSO)

Poglavje 4

Napovedovanje

Za napovedovanje proizvodnje smo uporabili različne modele strojnega učenja. Osredotočili smo se na dnevno napovedovanje, ki je najbolj uporabno na dnevnem trgu. Napovedni modeli, ki jih smo jih razvili, so prilagojeni za območje slovenske Primorske. Naš osnovni cilj je bil zgraditi model, ki je boljši od osnovnih modelov, opisanih v razdelku 4.2 in doseči skupno točnost normalizirane povprečne napake 3%. Pri tem smo uporabili 17 atributov, ki smo jih pridobili, izpeljali in izračunali. V podtjetju E3 zaenkrat še ne napovedujejo proizvodnje sončnih elektrarn, saj predstavljajo le 2% njihovega portfelja. Zaradi tega njihovega modela nismo mogli uporabiti za našo osnovo.

4.1 Izračun atributov

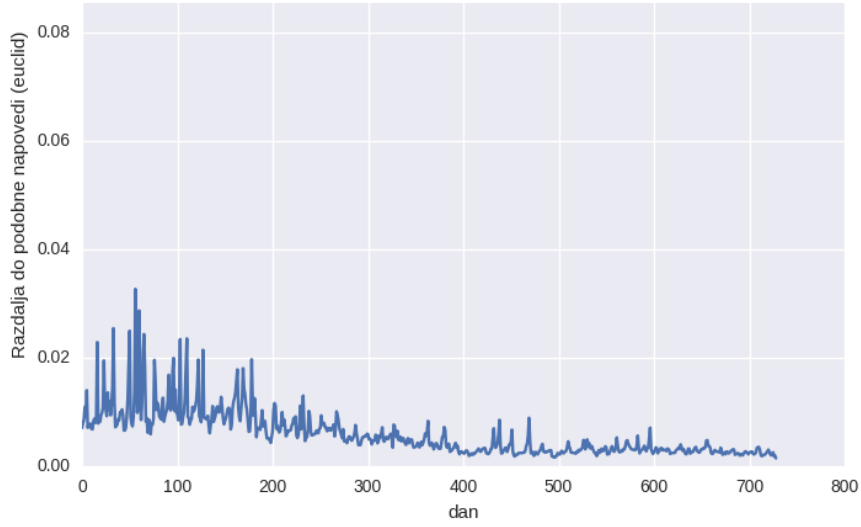
Za napovedovanje smo pridobili 13 atributov iz agencije ARSO, naštetih v tabeli 3.3. Uporabili smo vse attribute razen geopotenciala in indeksa za kopno in morje, saj bi imela ta dva atributa varianco 0. Geopotencial smo uporabili pri izračunu izpeljanega atributa indeksa sončnega sevanja G (enačba (4.1)). Poleg teh atributov smo izpeljali tudi svoje attribute.

- proizvodnja na podoben dan ob isti uri;
- proizvodnja na prejšnji dan ob isti uri;

- dolžina dneva;
- indeks sončnega sevanja G ;
- razmerje med neto kratkovalovnim sevanjem in indeksom sončnega sevanja.

Opravlka imamo s časovno odvisnimi podatki, zato smo dodali tudi proizvodnjo prejšnjega dne, ob isti uri in proizvodnjo podobnega dne, ob isti uri. Podoben dan smo izračunali na podlagi vremenskih napovedi z normalizirano neuteženo evklidsko razdaljo za vsako modelsko točko. Na sliki 4.1 je prikazana povprečna evklidska razdalja do najbližjega (najbolj podobnega) dne za vse modelske točke skupaj. Podoben dan iščemo vedno samo v preteklosti ($d_p < d_t$ kjer je d_p podoben dan in d_t trenuten dan), zato smo pričakovali, da se bo z večanjem časa (števíla dni) tudi manjšala razdalja, saj je večja verjetnost, da najdemo bolj podoben dan. Napovedi smo skušali izboljšati na podlagi podobnega dne določenega z uteženjem atributov. Uteži smo določili na podlagi izračunanih korelacij atributov s proizvodnjo (slike 4.4, 4.5 in 4.6). Pomagali smo si tudi s člankom [16], v katerem so obrazložene pomembnosti atributov. Atribute smo predhodno normalizirali, da so razdalje pred uteženjem enakovredne. Izkazalo se je, da je napoved z uteženim podobnim dnem slabša, zato smo pri napovedih uporabili neutežen podoben dan.

Poleg pridobljenih in zgodovinskih atributov smo izpeljali tudi nove. Dolžino dneva smo izračunali z urno razliko med sončnim zahodom in sončnim vzhodom. Ta je odvisna od vsake lokacije modelske točke, čeprav so zaradi majhnega območja razlike zelo majhne. S tem atributom smo želeli napovednim modelom pomagati, da dajo večjo pomembnost časovno podobnim dnem oziroma podobnim letnim časom. K atributom smo dodali tudi uro. Drug izpeljan atribut, indeks sončnega sevanja pri čistem nebu G (enačba (4.1)), smo, tako kot v literaturi [51, 52], uporabili tudi mi. Sončno sevanje je vrednost sončne energije prejete v določenem času na enoto površja Zemlje. Gre za matematično izračunane vrednosti, ki so odvisne od položaja v koordinatnem sistemu Zemlje (z zemljepisno dolžino in zemljepisno širino (ϕ)), časa

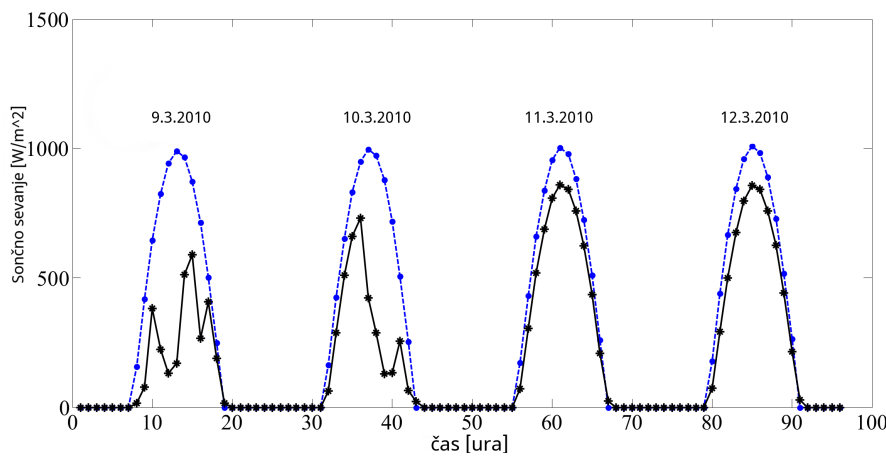


Slika 4.1: Povprečna evklidska razdalja do najbližjega dne za vse modelske točke skupaj.

(d za dan v letu), sončne deklinacije (kot med zveznico Zemlja-Sonce in ravnino ekvatorja označen kot δ), urnega kota (h_s) in solarne konstante (G_0). Solarna konstanta G_0 je določena na osnovi merjenja spektra sončnega sevanja na zunanjem robu atmosfere, torej nanjo ne vplivata vreme in atmosfera. Njena izmerjena vrednost je 1367 W/m^2 , izmerila pa jo je svetovna meteorološka organizacija [22]. Na sliki 4.2 vidimo primer sončnega sevanja štirih dni, za indeks sončnega sevanja z modro barvo in izmerjeno sončno sevanje na tleh s črno barvo. Iz razlike obeh krivulj lahko sklepamo kakšni so bili vremenski pogoji. V nabor atributov smo dodali še razmerje med neto kratkovalovnim sevanjem in indeksom sončnega sevanja (N_{ir}/G). Pri izračunih smo si pomagali s knjižnico pvlib-python [53].

$$G = G_0 \left(1 + 0.033 \cos\left(\frac{360d}{365}\right)\right) (\cos(\delta) \cos(h_s) \cos(\phi) + \sin(\delta) \sin(\phi)) \quad (4.1)$$

$$W_s = \sum_{i=0}^3 \frac{1}{d_i} \quad (4.2)$$



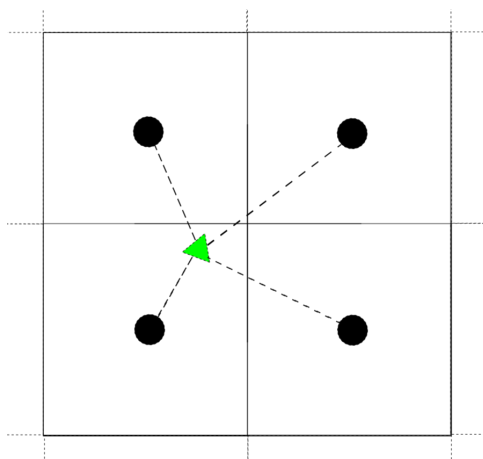
Slika 4.2: Prikaz izmerjenega sončnega sevanja s črno krivuljo in sončnega sevanja pri čistem nebu z modro krivuljo. (povzeto iz [51]).

$$W_i = \frac{\frac{1}{d_i}}{W_s} \quad (4.3)$$

$$\hat{X} = \sum_{i=0}^3 X * W_i \quad (4.4)$$

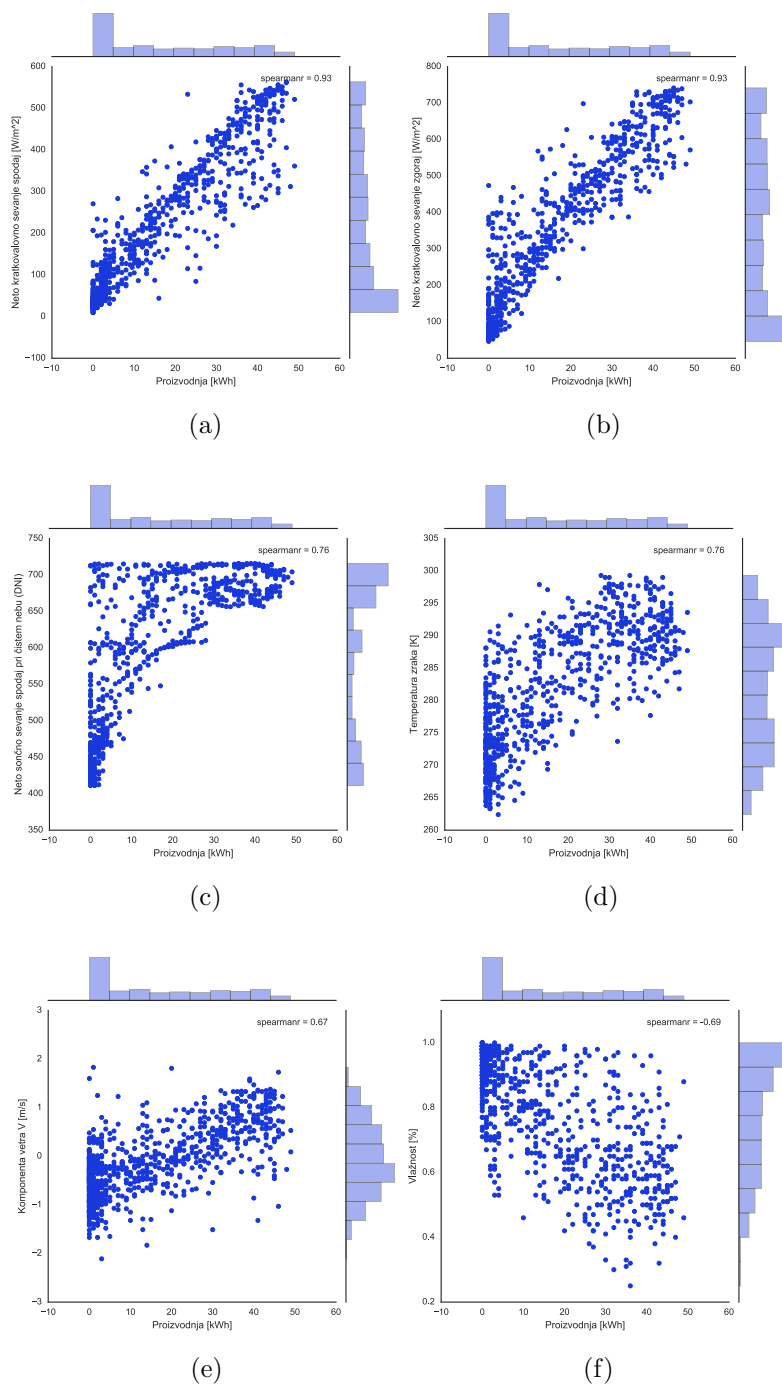
Zgradili smo tudi nove lokacijsko odvisne atribute. Namesto, da vzamemo najbližjo modelsko točko vremenskih napovedi, sedaj vzamemo štiri najbližje. Te napovedi smo utežili glede na oddaljenost elektrarne. Takega načina gradnje atributov nismo zasledili nikjer v literaturi, vendar smo ga vseeno preizkušali in primerjali s klasičnim načinom najbližje modelske točke. Uteži smo določili z ravninsko (bilinearno) interpolacijo tako, da smo vse štiri razdalje (d_i) normalizirali (4.2) in transformirali, da je njihova vsota 1 (4.3). Želeli smo doseči, da ima bližja modelska točka večji vpliv, bolj oddaljena pa manjši. Na sliki 4.3 so prikazane najbližje razdalje od elektrarne (zelen trikotnik) do modelskih točk (črni krogi). Nato smo vse modelske napovedi (X) zmnožili z utežmi in jih sešteli (4.4). Primerjava rezultatov napovedi z eno ali štirimi modelskimi točkami je prikazana v tabelah 5.1 in 5.2.

Preverili smo koreliranost vseh atributov z uporabo Spearmanovega korelacijskega koeficienta za vsako uro za dve elektrarni. Za vsako uro smo

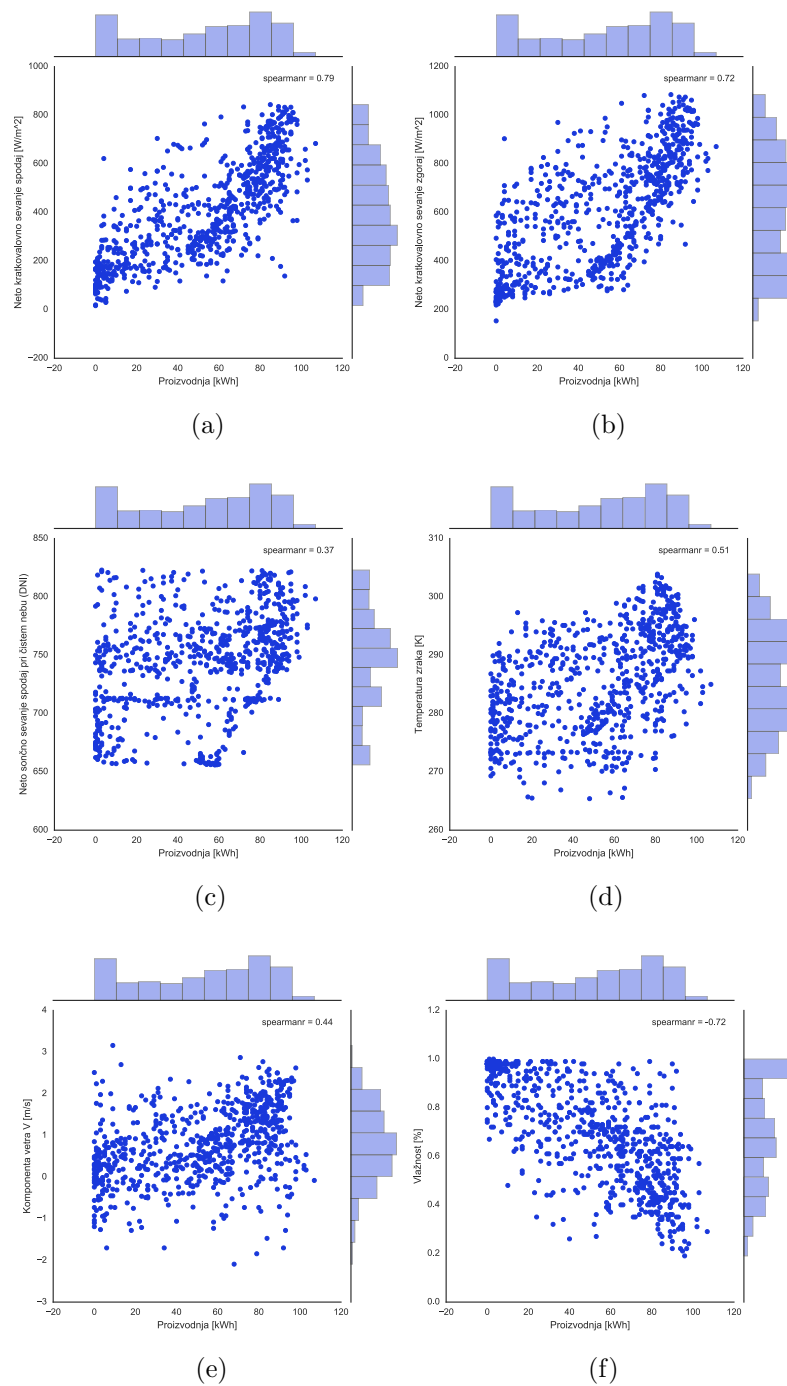


Slika 4.3: Primer lokacije elektrarne označene z zelenim trikotnikom med štirimi najbližjimi modelskimi točkami označenimi s črnimi krogi.

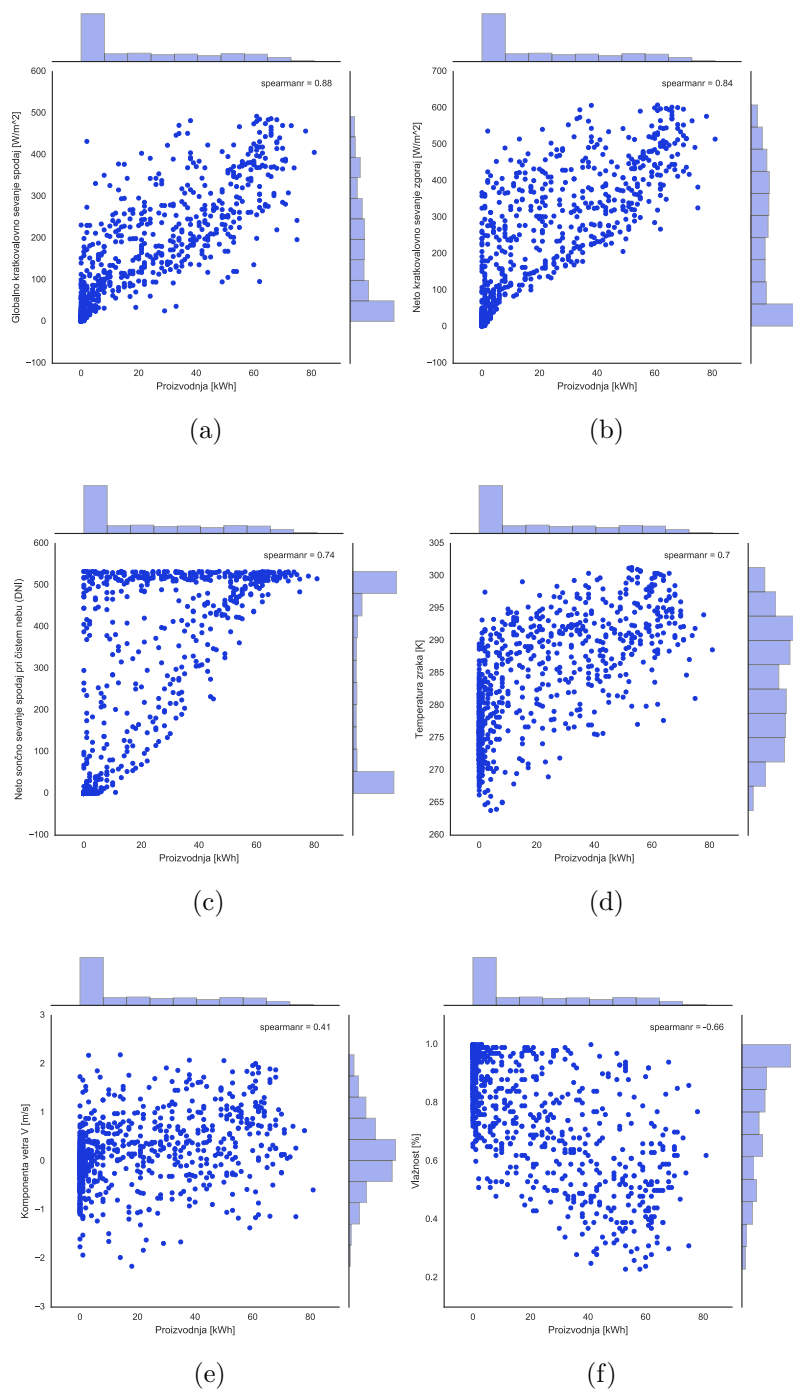
izračunali korelacijo med vsemi parnimi kombinacijami atributov. Na sliki 4.4, 4.5 in 4.6 so vidne nam najbolj zanimive in uporabne korelacije za ure 8:00, 12:00 in 16:00 za eno elektrarno. Vidimo, da ima atribut “kratkovalovno sevanje spodaj” največjo korelacijo s proizvodnjo. Zanimivo je, kako se korelacije s proizvodnjo čez dan (okrog ure 12:00) zmanjšajo. Še posebej je razvidno pri *sončnem sevanju pri čistem nebu*, kjer pade korelacija od jutra (8:00) do poldne (12:00) iz 0.76 na 0.37 (sliki 4.4.c in 4.5.c). Korelacija se zmanjša zaradi oblačnih in deževnih dni. Opazili smo nepričakovano veliko korelacijo (0.67) komponente vetra v , s proizvodnjo ob 8:00 uri (slika 4.4.e). Vlažnost ima pričakovano negativno korelacijo s proizvodnjo, saj vlaga odbija sončne žarke (slike 4.4.f, 4.5.f in 4.6.f). V članku [16] so predstavljena še druga opažanja, ki jih lahko razberemo iz teh korelacij. Za razliko od nas, so za mero koreliranosti uporabljali Pearsonov korelacijski koeficient. Za Spearmanovo korelacijo smo se odločili zato, ker primerja, če sta dve spremenljivki monotono povezani in lahko zazna razmerje tudi, če povezava ni linearna.



Slika 4.4: Korelacije med atributi ob 8:00 uri.



Slika 4.5: Korelacije med atributi ob 12:00 uri.



Slika 4.6: Korelacije med atributi ob 16:00 uri.

4.2 Modeliranje proizvodnje

$$\hat{y}_{t+\tau} = y_t \quad (4.5)$$

Naredili smo si dva osnovna modela (baseline), model prejšnjega dne in model podobnega dne. Model prejšnjega dne je v literaturi [19] znan tudi pod imenom “model vztrajnosti naslednjega dne” in v njem predpostavljamo, da bo se bodo vremenski pogoji ponovili tudi naslednji dan za določen časovni interval. V enačbi (4.5) je $\hat{y}_{t+\tau}$ napovedana vrednost spremenljivke y pri določenem času τ od začetnega časa t . Ker gre za model naslednjega dne, je vrednost τ 24 ur.

$$\tau = 24 d \quad (4.6)$$

Drugi intuitivno smiselni model, ki smo ga zgradili, je na podlagi podobnega dne. Podroben postopek pridobivanja najbolj podobnega dne je opisan v razdelku 4.1. Nova enačba je enaka prejšnji (4.5), le da je vrednost τ spremenjena. V enačbi (4.6) spremenljivka d predstavlja število dni od trenutne napovedi, do najbolj podobne napovedi. V članku [39] so predstavili podoben model imenovan “model podobnih instanc”. Z evklidsko razdaljo so izračunali razdalje do vseh podobnih dni, le da so to naredili na atributih po transformaciji PCA. Za učno množico so vzeli n najbolj podobnih dni, ostale pa so ignorirali. S tem modelom so napovedovali sončno sevanje in ne izhodne moči fotovoltaičnih modulov. Primerjave rezultatov z in brez transformacije PCA niso predstavili.

Ostali modeli, ki smo jih uporabljali so regresijska odločitvena drevesa, naključni gozdovi, k najbližjih sosedov (kNN) in linearna regresija z regularizacijo. Za vse smo uporabljali pythonovo knjižnico scikit-learn [18].

4.2.1 Testna metodologija

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.7)$$

$$MAE_z = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|; \Phi_z < 85 \quad (4.8)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (4.9)$$

$$nMAE = \frac{MAE}{MAX(PS)} \quad (4.10)$$

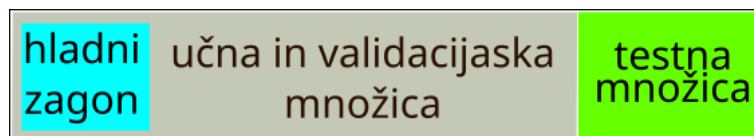
$$nMAE_z = \frac{MAE_z}{MAX(PS)} \quad (4.11)$$

$$nMBE = \frac{MBE}{MAX(PS)} \quad (4.12)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}; Y_i > 0 \quad (4.13)$$

Napako merimo s povprečno absolutno napako (MAE) (enačba (4.7)) in povprečno napako (MBE) (enačba (4.9)). V literaturi so nekateri avtorji uporabljali tudi povprečno absolutno procentno napako MAPE (enačba (4.13)), vendar zaradi težave, ki lahko pride pri deljenju z nič, lahko napaka limitira v neskončnost. Da smo se izognili temu problemu, smo dodali pogoj, da je izmerjena proizvodnja večja od 0. V literaturi ni omenjeno kako so se temu problemu izognili. Da smo napake lažje primerjali, smo jih normalizirali z močjo elektrarne (PS) [4.10, 4.12]. Napoved enega modela obsega 24 vrednosti, za katere smo izračunali obe napaki. Končno vrednost napake smo dobili s povprečjem vseh dnevnih napak učne oziroma testne množice (slika 4.7). Zaradi lažje primerjave z literaturo, smo izračunali tudi napako $nMAE_z$ pri pogoju ($\Phi_z < 85$) (enačba (4.11)), torej ko je sončni zenitni kot manjši od 85° , kar pomeni, da je takrat dan (sonce je že vzšlo in ni še zašlo).

Ker so naši podatki zelo vezani na čas, si jih moramo predstavljati kot podatkovni tok (data-stream). Modele gradimo za vsako elektrarno posebej in ker ima vsaka elektrarna lahko drugačen datum začetnega obratovanja, so datumi delitve podatkov na učno in testno množico nekoliko različni. Da

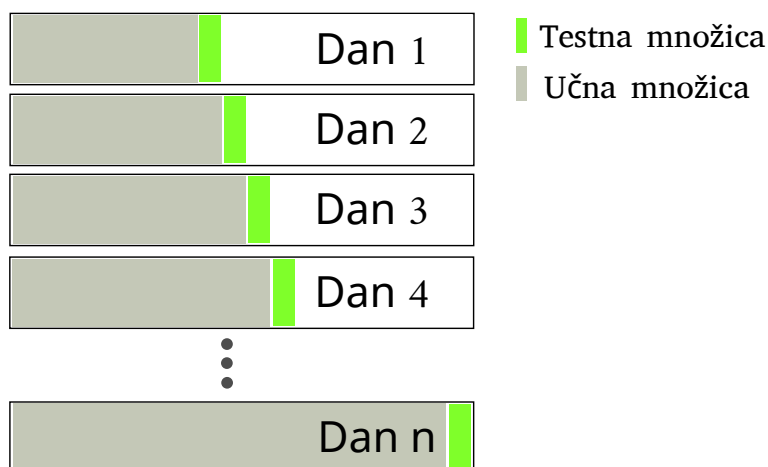


Slika 4.7: Prikaz razdelitve podatkov na učno in testno množico.

lahko smiselno poiščemo potrebne parametre za vhodni model, smo podatke razdelili na 2 dela, učno in testno množico, kot je razvidno iz slike 4.7. V učni množici uporabimo najprej prvih 50 dni za začetek učenja modela (hladni zagon), ostale dni pa model inkrementalno povečujemo pred učenjem in napovedujemo z določenim parametrom za vsak dan. Učna množica je velikosti 70% dni celotnega delovanja elektrarne. V članku [39] so za začetek učenja modela uporabljali množico velikosti največ 25 dni, vendar so imeli podatke le za eno leto. Za lažjo predstavo smo inkrementalno povečevanje množice ilustrirali na sliki 4.8. Če imamo na primer učno množico velikosti 200 veljavnih dni, to pomeni, da zgradimo 200-50 modelov, s katerimi napovedujemo naslednjih 24 ur. Testna množica predstavlja preostalih 30% dni celotnega delovanja elektrarne. Učenje z inkrementalnim povečevanjem učne množice izvajamo na učni in testni množici. Na testni množici gradimo modele s parametrom, pri katerem je učna množica dala najboljši rezultat z merjenjem napake MAE (enačba (4.7)). Rezultati testnih množic različnih modelov so tako med sabo primerljivi in na podlagi primerjave modela, ki da najmanjšo napako, lahko izberemo končni model.

4.2.2 Ocenjevanje atributov

Iz nabora atributov, predstavljenih v razdelku 4.1, smo skušali ugotoviti kombinacijo, ki bi dala najboljše rezultate. Najprej smo izločili attribute z varianco 0. Ker je število kombinacij vseh atributov preveliko, da bi za vsako kombinacijo preverili obnašanje modelov, smo skušali sami izbrati nekaj kombinacij. Attribute smo ocenili s pomočjo naključnih gozdov. Preverili



Slika 4.8: Inkrementalno povečevanje učne množice izvajano na množicah s slike 4.7.

smo kombinacije teh ocen tako, da smo testirali vse modele z atributom, ki je bil najvišje ocenjen, nato z dvema najvišje ocenjenima in tako naprej dokler nismo vključili vseh atributov. Izkazalo se je, da so rezultati boljši preden vključimo vse attribute. V tabeli 4.4 so razvidne ocene atributov z modelom naključnih gozdov. Poizkušali smo tudi transformacijo atributov z metodo PCA. V tabeli 4.5 so prikazane variance lastnih vektorjev (novih atributov). Iz tabele vidimo, da prva komponenta po transformaciji s PCA predstavlja 75% variance podatkov. Na podlagi teh vrednosti smo se odločili, da izberemo le tiste attribute, ki imajo varianco večjo od $1 * 10^{-3}$.

Za eno elektrarno, ki je začela obratovati pred 1. 1. 2011, smo zgradili ~ 784.000 modelov, da smo lahko preverili obnašanje z različnimi atributi in parametri. Celoten postopek smo ponovili še z zamenjavo vremenskih napovedi najbližje točke s povprečnimi vremenskimi napovedmi najbližjih štirih točk. Zaradi časovne omejitve smo ta postopek izvedli le na dveh elektrarnah, ki imajo podatke za celotno obdobje dveh let, za ostale elektrarne pa smo uporabili najboljše kombinacije atributov vrednotenih na dveh elektrarnah.

atribut	vrednost
podoben dan	0.5012
globalno kratkovalovno sevanje pri tleh	0.2420
neto kratkovalovno sevanje pri tleh	0.1037
ura	0.0267
prejšnji dan	0.0212
relativna vlažnost	0.0187
neto dolgovalovno sevanje pri tleh	0.0114
neto kratkovalovno sevanje na vrhu atmosfere	0.0110
u komponenta vetra	0.0102
v komponenta vetra	0.0100
indeks sončnega sevanja G	0.0093
oblačnost	0.0079
razmerje N_{ir}/G	0.0075
globalno dolgovalovno sevanje pri tleh	0.0072
temperatura zraka	0.0064
dolžina dneva	0.0025
količna padavin	0.0024

Tabela 4.4: Ocene atributov, pridobljene z naključnimi gozdovi.

PCA
0.7479
0.1888
0.0471
0.0114
0.0023
0.0008
0.0006
0.0003
0.0002
0.0001
7.1e-06
2.7e-06
2.1e-06
1.4e-06
6.6e-07
7.1e-08
2.7e-08

Tabela 4.5: Variance atributov, pridobljenih s transformacijo PCA.

Poglavje 5

Rezultati

V podjetju *E3* s trenutnimi napovedmi odjema in proizvodnje hidroelektrarn na tedenskem trgu dosegajo v letnem profilu napako do 3%. Ker je to trenutno njihov pričakovani rezultat, s katerim lahko pozitivno poslujejo, bi tudi napovedi sončnih elektrarn z napako do 0.03 nMBE prinesle dodatno točnost v njihov portfelj napovedi. Na sliki A.1 in 5.1 je graf povprečne napake nMAE, na sliki A.2 in 5.2 pa je graf povprečne napake nMBE napovedi naključnih gozdov, za vsako uro ene elektrarne, z uporabo vremenskih napovedi iz ene oziroma štirih modelskih točk. Vsaka krivulja predstavlja določen nabor atributov. Postopek določanja nabora je opisan v razdelku 4.2.2. Iz grafov smo videli obnašanje napak in smiselno določili optimalen nabor atributov. Na sliki 5.3 so prikazane dnevne napake nMAE ene elektrarne na testni množici. Napaka se proti koncu leta niža. Razlog je verjetno ta, da je dan krajši in da se učna množica veča. V tabelah 5.1 in 5.2 so razvidne povprečne napake vseh modelov za najbližjo modelsko točko in za povprečje štirih modelskih točk. V dodatku smo v tabelah A.1 in A.2 dodali podrobnejše rezultate. V tabelah 5.1 in 5.2 število dni pod 3% (št. dni pod 3%) pove, koliko dni je bila v testni množici napaka nMAE manjša od 3%. Največja testna množica je velikosti 219 dni. V obeh tabelah je rezultat modela prejšnjega in podobnega dne enak. Tak rezultat smo tudi pričakovali, vendar ko smo pri iskanju podobnega dne attribute utežili, se je model po-

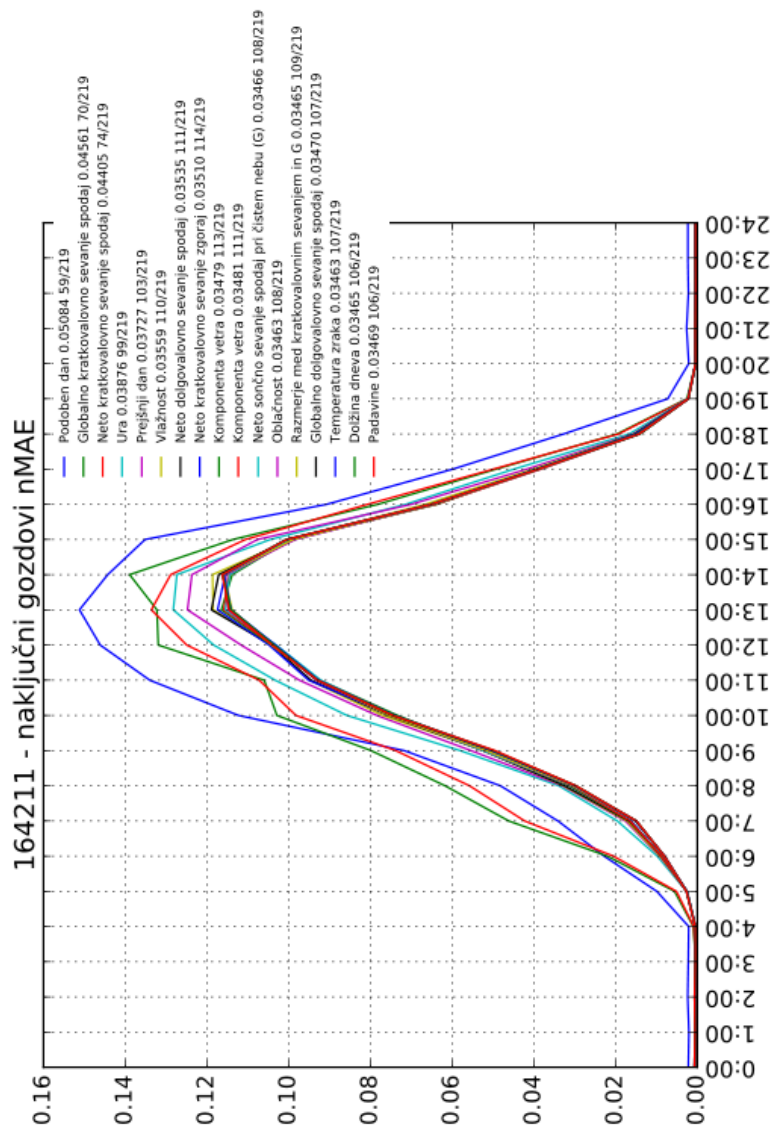
dobnega dne spremenil.

V tabeli 5.3 smo predstavili rezultate napovedi naključnih gozdov enkratnega, tedenskega in vsakodnevnega inkrementalnega povečevanja učne množice. Najboljši rezultat smo dobili z vsakodnevnim povečevanjem učne množice. Razlike v napovedni točnosti so relativno majhne, vendar je zaradi majhne časovne omejitve uporaba vsakodnevnega povečevanja množice smiselna. Časovna omejitev je osem ur od generirane modelske napovedi do časa, ki ga mora podjetje E3 oddati na borzi.

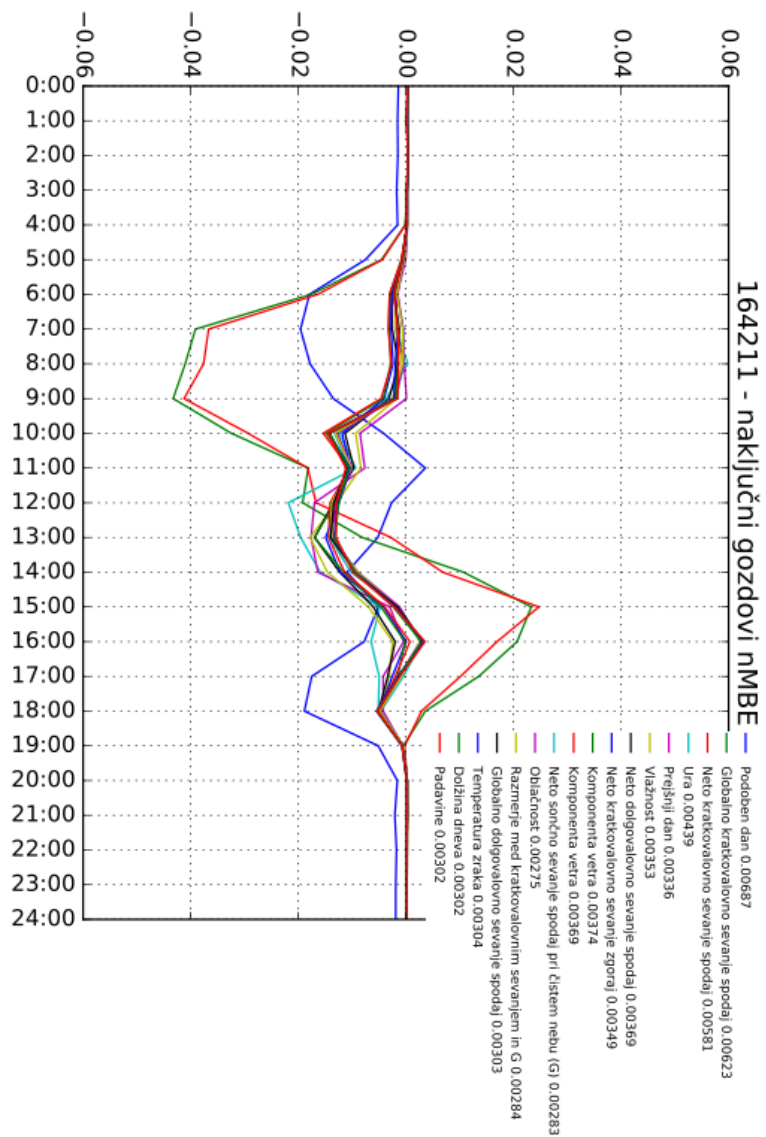
Zaradi težav z nekonsistentno uporabo metrik napak v tuji literaturi, smo se osredotočili le na 2 članka, v katerih je viden najboljši rezultat in ga lahko primerjamo z našimi rezultati. V tabeli 5.4 je rezultat iz članka [19] označen z NAM_z . V tem članku so rezultati analize le dveh elektrarn iz jugozahodne Amerike, v katerem so uporabljali neutruženo povprečje vremenskih napovedi. Z oznako $PVSAT$ smo označili rezultate iz članka [33], v katerem so rezultati analize desetih elektrarn iz Kanade. Z indeksom z smo označili rezultate, ko je sončni zenitni kot manjši od 85° . Iz obeh člankov so rezultati normalizirani z maksimalno nazivno proizvodnjo elektrarn. Ker je nazivna proizvodnja napisana s strani proizvajalca in v praksi ni nikoli dosežena, se moramo zavedati, da so rezultati teh dveh člankov lahko zaradi tega razloga boljši v primerjavi z našo metriko napak. Prav tako iz teh člankov ni jasno razvidno, ali so za napovedi uporabljali inkrementalno povečevanje učne množice, kot smo ga v našem delu.

5.1 Sklepne ugotovitve

Ugotovili smo, da najboljše napovedi dobimo z modelom naključnih gozdov, če pri tem uporabljamo le prvih 12 atributov, urejenih z oceno atributov iz naključnih gozdov (tabela 4.4). Z interpolacijo napovedi modelskih točk smo rezultate dodatno še izboljšali. Zanimivo je, da smo po transformaciji atributov s PCA dobili z modelom naključnih gozdov slabšo napoved od najbližjih sosedov (kNN). Brez transformacije PCA pa se je model kNN precej slabše iz-



Slika 5.1: Graf povprečnih napak napovedi nMAE ene elektrarne za vsako uro z napovedjo iz štirih modelskih točk.



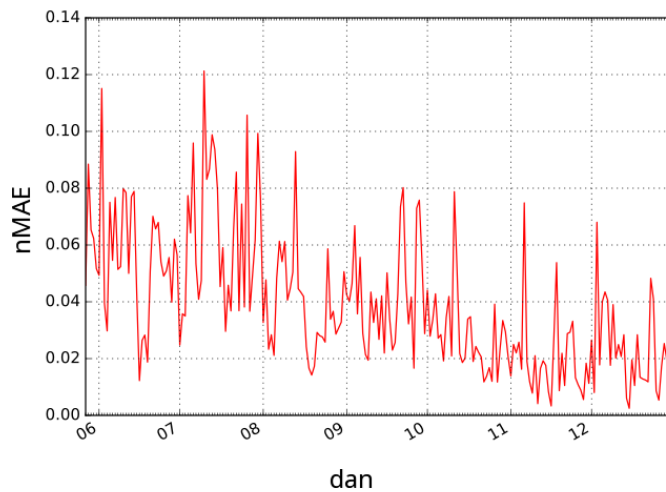
Slika 5.2: Graf povprečnih napak napovedi nMBE ene elektrarne za vsako uro z napovedjo iz štirih modelskih točk.

nabor atributov	model	št. dni pod 3%	nMBE	nMAE	nMAE _z	MAPE	nMBE _{std}
orig.	kNN	138	0.00154	0.04089	0.08880	0.00519	0.03195
orig.	naključni gozdovi	199	0.00270	0.03559	0.07704	0.01713	0.02971
orig.	linearna regresija	102	0.00233	0.04486	0.08671	0.00453	0.03102
orig.	regresijska drevesa	166	0.00239	0.03948	0.08519	0.02741	0.03038
orig.	prejšnji dan	154	0.00011	0.05086	0.11105	0.00567	0.05856
orig.	podoben dan	114	0.00955	0.04765	0.10304	0.06906	0.03813
PCA	kNN	132	0.00288	0.04252	0.09195	0.00279	0.03246
PCA	naključni gozdovi	121	0.00339	0.04301	0.09286	0.00763	0.03259
PCA	linearna regresija	44	0.00250	0.05386	0.10547	0.00502	0.03460
PCA	regresijska drevesa	112	0.00344	0.04532	0.09694	0.00885	0.03443

Tabela 5.1: Rezultati napovednih modelov za najbližjo modelsko točko.

nabor atributov	model	št. dni pod 3%	nMBE	nMAE	nMAE _z	MAPE	nMBE _{std}
orig.	kNN	141	0.00166	0.04037	0.08770	0.00390	0.03132
orig.	naključni gozdovi	201	0.00242	0.03538	0.07665	0.01581	0.02889
orig.	linearna regresija	100	0.00263	0.04471	0.08633	0.00523	0.03079
orig.	regresijska drevesa	169	0.00244	0.03960	0.08546	0.02528	0.03136
orig.	prejšnji dan	154	0.00011	0.05086	0.11105	0.00567	0.05856
orig.	podoben dan	114	0.00955	0.04765	0.10304	0.06906	0.03813
PCA	kNN	134	0.00320	0.04230	0.09147	0.00265	0.03240
PCA	naključni gozdovi	123	0.00344	0.04278	0.09244	0.00779	0.03210
PCA	linearna regresija	42	0.00298	0.05369	0.10478	0.00462	0.03436
PCA	regresijska drevesa	106	0.00403	0.04524	0.09690	0.00822	0.03439

Tabela 5.2: Rezultati napovednih modelov za povprečje modelskih točk.



Slika 5.3: Dnevne napake nMAE ene elektrarne na testni množici.

gradnja modela	nMAE	nMBE
enkrat	0.03561636	0.0032618
tedensko	0.03506643	0.00325115
vsak dan	0.03462877	0.00275618

Tabela 5.3: Primerjava gradnje modela na različnih časovnih intervalih za eno elektrarno.

metrika	PVSAT	NAM _z	naš model
nMAE	3.66	-	3.54
nMAE _z	-	7.8	7.67
nMBE	0.433	-	0.24
nMBE _z	-	0.25	0.47

Tabela 5.4: Primerjava rezultatov s tujimi raziskavami.

kazal. Z našim modelom nismo dosegli željene povprečne napake nMAE pod 3%, vendar smo se ji precej približali. V kontekstu Slovenskega območja električnega sistema napaka nNMAE 3.5% (ki smo jo dosegli) relativno ni velika, saj sončne elektrarne predstavljajo le $\sim 1\%$ celotne proizvodnje v Sloveniji. Zelo dobre rezultate smo dobili glede na napako nMBE. Iz te napake vidimo, da naš model napovedovanja električne energije ne sistematično napoveduje premajhne ali prevelike proizvodnje. Kljub temu, da je MBE pri metodah KNN in “prejšnji dan” najnižji, sta metodi manj primerni za praktično uporabo zaradi tudi do 100% višje standardne deviacije. To pomeni, da v praksi zaradi prevelike napake večkrat presežejo zastavljeno mejo napake 3% (tabeli 5.1 in 5.2). Ta napaka je bolj relevantna za trgovanje na borzi električne energije, ker predstavlja dejansko odstopanje vseh elektrarn skupaj. Razlog za zelo majhno napako je v velikem številu opazovanih elektrarn, saj se odstopanja v večji meri izničijo. Glede na to, da je vreme zelo težko napovedljivo in da so naše napovedi zelo odvisne od vremenskih, menimo, da je rezultat zelo dober. Čeprav so rezultati zelo težko primerljivi med sabo, smo dobili celo boljše rezultate od tujih avtorjev. Za pravo primerjavo ocenjujemo, da bi modele morali testirati na enakih lokacijah z enakimi kriteriji.

Poglavje 6

Zaključek

V magistrskem delu smo se osredotočili na kratkoročno napovedovanje dnevne proizvodnje sončnih elektrarn na področju slovenske Primorske. Cilj magistrskega dela je bil raziskati področje vpliva napovedi sončnih elektrarn in razviti čim bolj točen napovedni model. Ker gre za elektrarne na slovenskem ozemlju, smo opisali tudi trgovanje na slovenskem elektro energetskega trgu. Do problema smo pristopili sistematično in analitično. Skozi različne faze smo modele gradili do dosega zadovoljivih rezultatov. Imeli smo veliko birokratskih težav pri pridobivanju podatkov in nam brez poznanstev najverjetneje sploh ne bi uspelo. Podatki, ki smo jih dobili, so bili nepričakovano slabo organizirani in neurejeni. To lahko tudi pojasni, zakaj je tako malo objavljenih raziskav na temo napovedovanja proizvodnje sončnih elektrarn. Za napoved proizvodnje smo uporabljali vremenske napovedi, zgodovinske, lokacijske in izpeljane podatke. Podatkov iz vremenskih postaj nismo uporabljali, ker je mreža vremenskih postaj preredka. Na Primorskem sta le 2 glavni vremenski postaji in štiri klimatološke, ostale so padavinske. Postaje imajo različne nepopolne nabore senzorjev, kar prinaša dodatne težave pri vključitvi v model.

V Nemčiji so najbolj znana podjetja Meteogroup¹, Energymeteo² in Ener-

¹<http://www.meteogroup.com/>

²<https://www.energymeteo.com/en/index.php>

cast³, ki napovedujejo proizvodnjo iz sončnih elektrarn. Za napovedi uporabljajo vremenske napovedi iz numeričnih modelov, podatke vremenskih postaj in satelitskih slik. Obstajajo tudi podjetja, kot je na primer Stedysun⁴, ki napovedujejo tudi manjše časovne intervale (do 30 minut) s pomočjo slik zajetih iz kamer, ki so nameščene na elektrarni. Pri nas takšnih podjetji še ni.

Naš napovedni model se je izkazal kot robusten, saj sta napaki nMBE in NMAE nizki v vseh sezonskih obdobjih ter v različnih vremenskih pogojih. Verjamemo, da bi lahko imeli tudi natančnejšo napoved proizvodnje, vendar bi potrebovali daljše meritve opazovanj, ki bi bile izvedene s senzorji (kamera ali piranometer), postavljenimi neposredno na vsaki posamezni elektrarni.

Za prostor primorske regije Slovenije smo dosegli velik napredek, ki je pomemben predvsem za podjetji, za kateri se je raziskava delala. Naše delo ima velik pomen za modernizacijo elektroenergetskih sistemov, cenejših obratovalnih stroškov in posredno pripomore k prijaznejšemu okolju. Raziskava lahko služi kot poslovna ideja za novo podjetje prodaje dnevniških napovedi proizvodnje električne energije. V Sloveniji takega komercialnega produkta še ni, ampak imajo podjetja za ta namen interne ekipe. Raziskava je uporabna tudi za širšo uporabo. Metode čiščenja in validiranja podatkov lahko uporabimo tudi na ostalih podatkih elektrarn OVE (na primer vetrne elektrarne), metoda napovedovanja pa je v večji meri lahko primerna tudi za ostale države in regije.

6.1 Nadaljne delo

Ker bodo kmalu (vrjetno že konec leta 2016) na voljo preostali podatki vremenskih napovedi in ker se število napak v proizvodnji manjša, bi bilo smiselno testirati podatke tudi za kasnejša tri leta. Tako bi dobili boljši pregled nad celoletno napovedjo. Trenutno so v model vključeni le napovedni atributi

³<http://www.enercast.de/en/>

⁴<http://steady-sun.com/>

vremenskih napovedi, možnost pa je, da bi vključili tudi podatke izmerjene iz vremenskih postaj in satelitskih slik. V tem kontekstu bi lahko preverili odvisnost napake napovedi proizvodnje od napake vremenske napovedi. Zelo pomembno bi bilo vključiti v napovedovanje izredne dogodke, kot je na primer Lunin ali Sončev mrk in vključevanje vpliva snega. Po sneženju lahko namreč sneg ostane na panelih še nekaj dni.

Raziskali bi lahko vpliv podobnega dne oziroma nabora več podobnih dni. Sami smo smiselno utežili atribute, vendar smo dobili slabše rezultate. Potrebno bi bilo opraviti podrobnejšo analizo uteži, saj obstaja možnost, da bi s tem napovedi še izboljšali. Poiskati bi morali tudi informacijo o tem, koliko podobnih dni je smiselno vključiti v model. V naši raziskavi smo namreč vključili le enega.

Smiselno bi bilo izračunati, koliko modelskih točk je potrebnih, da so napovedi še vedno zadovoljive. S tem bi lahko zmanjšali odkupno ceno vremenskih napovedi in povečali prihranke pri podjetju, kot je na primer E3. Smiselna bi bila tudi primerjava z statističnimi modeli časovnih vrst, kjer bi uporabljali le izmerjene podatke za vhodni model napovedovanja 2.2 (na primer ARMAX).

Potrebno bi bilo izdelati še končni sistem in ga integrirati v realni čas napovedovanja, da bi lahko začeli uporabljati napovedi v praktične namene, opisane v 1.1. Sistem napovedovanja bi bilo potrebno tudi razširiti na vetrne in male hidroelektrarne, da lahko pokrijemo celoten spekter OVE.

Literatura

- [1] K. Sredenšek, "Vpliv naklona fotonapetostnih modulov na proizvodnjo električne energije," 2014. Diplomsko delo, Fakulteta za elektrotehniko, računalništvo in informatiko (FERI).
- [2] "Podporna shema proizvodnje električne energije iz OVE in SPTE, ministrstvo za infrastrukturo." Dostopno na: <http://www.energetika-portal.si/podrocja/energetika/podporna-shema-ove-in-spte/>, 2016. (pridobljeno: 18.8.2016).
- [3] "The state of renewable energies in europe." Dostopno na: <https://www.borzen.si/Portals/0/SL/Splo%C5%A1no/BROSURA%20web.pdf>, 2015. (pridobljeno: 1.10.2016).
- [4] "The state of renewable energies in europe." www.energies-renouvelables.org/observ-er/stat_baro/barobilan/barobilan14-gb.pdf, 2014. (pridobljeno: 24.8.2016).
- [5] "Smernice EU." Dostopno na: <https://ec.europa.eu/energy/en/topics/renewable-energy>, 2016. (pridobljeno: 23.5.2016).
- [6] M. B. Rajer, "Prihodnost podpor za proizvodnjo elektrike." Dostopno na: <http://beta.finance-on.net/files/2016-04-25/4-3-Predstavitev-Rajer-Portoroz.pdf>, 2016. (pridobljeno: 1.10.2016).
- [7] V. Habjan, "Pametna omrežja v praksi." Dostopno na: <http://www.nas-stik.si/1/Novice/novice/tabid/87/ID/2449/>

- Dobre-izkusnje-virtualne-elektrarne-skupine-Elektra-Ljubljana.aspx, 2014. (pridobljeno: 1.10.2016).
- [8] M. Svetina and P. Nemček, “Virtual power plants and utilities,” *Metering International*, no. 3, p. 133, 2011.
- [9] Y. Ren, P. Suganthan, and N. Srikanth, “Ensemble methods for wind and solar power forecasting—a state-of-the-art review,” *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 82–91, 2015.
- [10] “Uradni list republike slovenije, pravila za delovanje trga z električno energijo.” Dostopno na: <https://www.uradni-list.si/1/content?id=30724>, 2016. (pridobljeno: 25.10.2016).
- [11] D. D. Paravan, “Osnove trga z električno energijo (gen-i).” Dostopno na: http://www.powerlab.uni-mb.si/Slo/download/Literatura/Paravan-Osnove_trga_z_elektricno_energijo.pdf, 2016. (pridobljeno: 17.7.2016).
- [12] J. A. Podkoritnik, “Trg električne energije v sloveniji,” 2012. Diplomsko delo visokošolskega strokovnega študija Organizacija in management kadrovskih in izobraževalnih procesov.
- [13] “Regionalna energetska borza.” Dostopno na: <http://www.bsp-southpool.com/trgovanje-znotraj-dneva.html>, 2016. (pridobljeno: 17.7.2016).
- [14] “Holding slovenske elektrarne (hse).” Dostopno na: <http://www.hse.si/si/hse/predstavitev>, 2016. (pridobljeno: 15.7.2016).
- [15] “Crisp-DM.” Dostopno na: https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining, 2016. (pridobljeno: 29.6.2016).

-
- [16] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in *Smart Grid Communications (SmartGridComm)*, pp. 528–533, IEEE, 2011.
- [17] Y. Gala, Á. Fernández, J. Díaz, and J. R. Dorronsoro, "Hybrid machine learning forecasting of solar radiation values," *Neurocomputing*, vol. 176, pp. 48–59, 2016.
- [18] "Knjižnica scikit-learn." Dostopno na: <http://scikit-learn.org/stable/>, 2014. (pridobljeno: 25.8.2016).
- [19] D. P. Larson, L. Nonnenmacher, and C. F. Coimbra, "Day-ahead forecasting of solar power output from photovoltaic plants in the american southwest," *Renewable Energy*, vol. 91, pp. 11–20, 2016.
- [20] T. V. Žiga Kokalj, "Brošura center za daljinsko zaznavanje." Dostopno na: http://www.space.si/wp-content/uploads/2015/09/brosura_CDZ_SLO_web_201601.pdf, 2016. (pridobljeno: 24.6.2016).
- [21] "Mikro sončna elektrarna UM FERI." Dostopno na: <http://elektrarna.feri.uni-mb.si/>, 2016. Fakulteta za elektrotehniko, računalništvo in informatiko (FERI), (pridobljeno: 13.9.2016).
- [22] A. Tomažič, "Primerjava delovnih karakteristik različnih fotonapetostnih modulov pri enakih pogojih sončnega sevanja," 2010. Diplomsko delo, Fakulteta za elektrotehniko, računalništvo in informatiko (FERI).
- [23] S. Seme, *Optimalno sledenje fotonapetostnega sistema soncu ob upoštevanju izgub pogonskega sklopa*. PhD thesis, Fakulteta za elektrotehniko, računalništvo in informatiko (FERI), 2011.
- [24] K. Sheuermann, "Just how big is a 2 kW photovoltaic system?," *HOME ENERGY*, vol. 20, no. 1, pp. 24–30, 2003.
- [25] "Laboratorij za fotovoltaike in optoelektroniko." Dostopno na: <http://lpvo.fe.uni-lj.si/raziskave/fotovoltaika/>

- napovedovanje-energijskega-izplena-soncnih-elektrarn/, 2016. (pridobljeno: 13.8.2016).
- [26] G. Vozelj, "Kratkoročno napovedovanje proizvodnje električne energije sončnih elektrarn in malih hidroelektrarn," 2016. Magistrsko delo, Univerza v Ljubljani, Fakulteta za elektrotehniko (FE).
- [27] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.
- [28] D. Masa-Bote, M. Castillo-Cagigal, E. Matallanas, E. Caamaño-Martín, A. Gutiérrez, F. Monasterio-Huelín, and J. Jiménez-Leube, "Improving photovoltaics grid integration through short time forecasting and self-consumption," *Applied Energy*, vol. 125, pp. 103–113, 2014.
- [29] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 2, no. 1, pp. 2–10, 2009.
- [30] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, "Regional pv power prediction for improved grid integration," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 7, pp. 757–771, 2011.
- [31] J. G. d. S. F. Junior, T. Oozeki, H. Ohtake, K.-i. Shimose, T. Takashima, and K. Ogimoto, "Regional forecasts and smoothing effect of photovoltaic power generation in japan: An approach with principal component analysis," *Renewable Energy*, vol. 68, pp. 403–413, 2014.
- [32] J. G. d. S. Fonseca Junior, T. Oozeki, H. Ohtake, T. Takashima, and K. Ogimoto, "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods," *Progress in Photovoltaics: Research and Applications*, vol. 23, no. 10, pp. 1203–1218, 2015.

-
- [33] G. K. Sophie Pelland, George Galanis, “Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model,” *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 3, pp. 284–296, 2013.
 - [34] M. Zamo, O. Mestre, P. Arbogast, and O. Pannekoucke, “A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part i: Deterministic forecast of hourly production,” *Solar Energy*, vol. 105, pp. 792–803, 2014.
 - [35] M. Zamo, O. Mestre, P. Arbogast, and O. Pannekoucke, “A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. part ii: Probabilistic forecast of daily production,” *Solar Energy*, vol. 105, pp. 804–816, 2014.
 - [36] M. P. Almeida, O. Perpiñán, and L. Narvarte, “Pv power forecast using a nonparametric pv model,” *Solar Energy*, vol. 115, pp. 354–368, 2015.
 - [37] A. Mellit and A. M. Pavan, “A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected pv plant at trieste, italy,” *Solar Energy*, vol. 84, no. 5, pp. 807–821, 2010.
 - [38] Y. Li, Y. Su, and L. Shu, “An ARMAX model for forecasting the power output of a grid connected photovoltaic system,” *Renewable Energy*, vol. 66, pp. 78–89, 2014.
 - [39] R. Urraca, J. Antonanzas, M. Alia-Martinez, F. Martinez-de Pison, and F. Antonanzas-Torres, “Smart baseline models for solar irradiation forecasting,” *Energy Conversion and Management*, vol. 108, pp. 539–548, 2016.
 - [40] C. Renno, F. Petito, and A. Gatto, “Artificial neural network models for predicting the solar radiation as input of a concentrating photovoltaic

- system,” *Energy Conversion and Management*, vol. 106, pp. 999–1012, 2015.
- [41] M. Kozelj, “Pretvarjanje med različnimi koordinatnimi sistemi.” Dostopno na: <http://www.mkx.si/geoconv/>, 2016. (pridobljeno: 7.10.2016).
- [42] M. Rihtar, “Pretvornik koordinat.” Dostopno na: <http://geocoordinateconverter.tk/>, 2016. (pridobljeno: 12.4.2016).
- [43] M. Mataija, M. Pogarčič, and I. Pogarčič, “Helmert transformation of reference coordinating systems for geodesic purposes in local frames,” *Procedia Engineering*, vol. 69, pp. 168–176, 2014.
- [44] “Openstreetmap.” Dostopno na: <http://openstreetmap.org>, 2016. (pridobljeno: 27.4.2016).
- [45] E. Akarslan and F. O. Hocaoglu, “A novel adaptive approach for hourly solar radiation forecasting,” *Renewable Energy*, vol. 87, pp. 628–633, 2016.
- [46] “Standardizirani merilni in obračunski podatki, SODO.” Dostopno na: https://www.sodo.si/_files/903/Standardizirani_merilni_in_obracunski_podatki_V2_mar_2015.pdf, 2014. (pridobljeno: 19.5.2016).
- [47] “Knjiznjica scipy.” Dostopno na: <http://docs.scipy.org/>, 2016. (pridobljeno: 1.6.2016).
- [48] “Agencija republike Slovenije za okolje (arso).” Dostopno na: <http://www.arso.gov.si/>, 2016. (pridobljeno: 4.10.2016).
- [49] “Napovedni model Aladin-SI.” Dostopno na: <http://meteo.arso.gov.si/met/sl/faq/>, 2016. (pridobljeno: 19.7.2016).
- [50] J. Cedilnik, “Uporaba modela Aladin v klimatološke namene.” Dostopno na: http://www.fgg.uni-lj.si/sugg/referati/2005/SZGG_05_Cedilnik.pdf, 2005. (pridobljeno: 3.8.2016).

-
- [51] F. Wang, Z. Mi, S. Su, and H. Zhao, “Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters,” *Energies*, vol. 5, no. 5, pp. 1355–1370, 2012.
 - [52] Y. Ren, P. Suganthan, and N. Srikanth, “Ensemble methods for wind and solar power forecasting—a state-of-the-art review,” *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 82–91, 2015.
 - [53] W. Holmgren, “PVLIB Python.” Dostopno na: <http://pvlib-python.readthedocs.io/en/latest/>, 2016. (pridobljeno: 25.8.2016).

Dodatek A

Podrobnejši rezultati napovedi

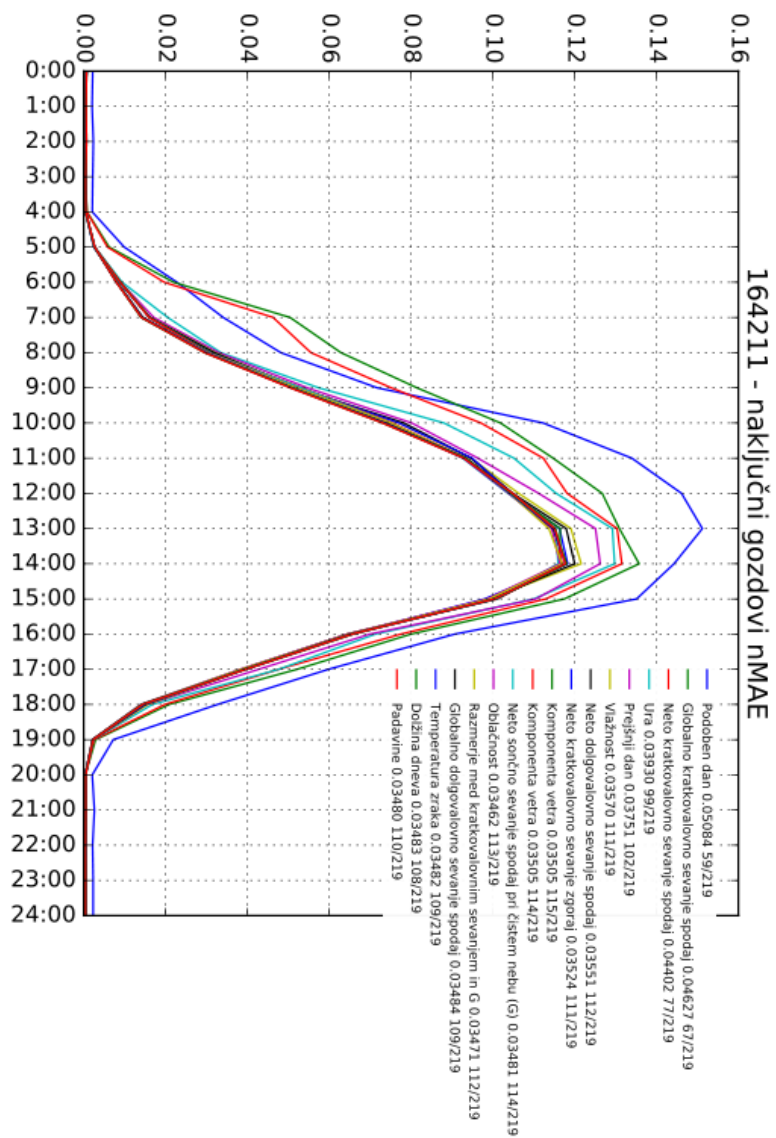
V tabelah A.1 in A.2 so predstavljeni podrobnejši rezultati napovedi proizvodnje sončnih elektrarn, pri katerih so bile uporabljene značilke najbližje modelske točke. Na grafih A.1 in A.2 so prikazane povprečne napake napovedi različnih naborov značilk najbližje modelske točke za metriki nMAE in nMBE ene elektrarne za vsako uro. Napovedni model temelji na naključnih gozdovih.

nabor atributov	model	nMBE _{min}	nMBE _{max}	nMBE _{std}	nMAE _{min}	nMAE _{max}	nMAE _{std}
orig.	kNN	0.00000	0.25242	0.03195	0.00531	0.25242	0.02341
orig.	naključni gozdovi	0.00007	0.22980	0.02971	0.00527	0.22980	0.02339
orig.	linearna regresija	0.00000	0.21218	0.03102	0.01004	0.21750	0.02229
orig.	regresijska drevesa	0.00001	0.22189	0.03038	0.00726	0.22189	0.02323
orig.	prejšnji dan	0.00000	0.28247	0.05856	0.00000	0.28247	0.04183
orig.	podoben dan	0.00002	0.29736	0.03813	0.00229	0.29736	0.03065
PCA	kNN	0.00000	0.18384	0.03246	0.00625	0.18384	0.02288
PCA	naključni gozdovi	0.00010	0.18047	0.03259	0.00431	0.18047	0.02312
PCA	linearna regresija	0.00012	0.19498	0.03460	0.01308	0.19521	0.02188
PCA	regresijska drevesa	0.00012	0.19242	0.03443	0.00494	0.19242	0.02360

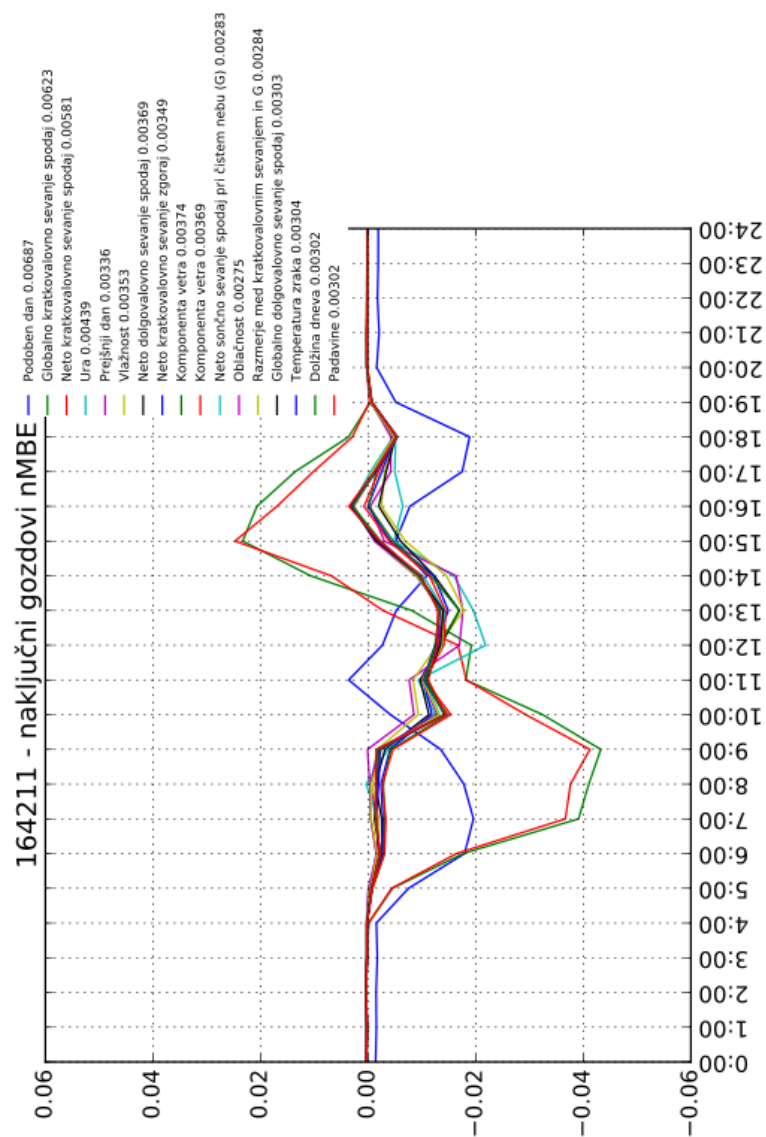
Tabela A.1: Rezultati napovednih modelov za najbližjo modelsko točko.

nabor atributov	model	nMBE _{min}	nMBE _{max}	nMBE _{std}	nMAE _{min}	nMAE _{max}	nMAE _{std}
orig.	kNN	0.00005	0.25633	0.03132	0.00534	0.25633	0.02294
orig.	naključni gozdovi	0.00020	0.21722	0.02889	0.00613	0.21722	0.02266
orig.	linearna regresija	0.00002	0.21747	0.03079	0.01000	0.22166	0.02222
orig.	regresijska drevesa	0.00004	0.26070	0.03136	0.00730	0.26332	0.02487
orig.	prejšnji dan	0.00000	0.28247	0.05856	0.00000	0.28247	0.04183
orig.	podoben dan	0.00002	0.29736	0.03813	0.00229	0.29736	0.03065
PCA	kNN	0.00000	0.18279	0.03240	0.00595	0.18279	0.02292
PCA	naključni gozdovi	0.00005	0.19637	0.03210	0.00436	0.19637	0.02309
PCA	linearna regresija	0.00001	0.20117	0.03436	0.01300	0.20121	0.02195
PCA	regresijska drevesa	0.00008	0.20013	0.03439	0.00485	0.20013	0.02389

Tabela A.2: Rezultati napovednih modelov za povprečje modelskih točk.



Slika A.1: Graf povprečnih napak napovedi nMAE ene elektrarne za vsako uro z napovedjo iz ene modelske točke.



Slika A.2: Graf povprečnih napak napovedi nMBE ene elektrarne za vsako uro z napovedjo iz ene modelske točke.